

7 Stereoskope HD-Produktion

NIKOLAUS HOTTONG / PAUL WALTER (HRSG.)

HOTTONG / WALTER (HRSG.)

Das Forschungsprojekt Stereoskope HD-Produktion an der Hochschule Furtwangen beleuchtet das Feld der 3D-Filmproduktion unter den aktuellen Randbedingungen von High Definition Video mit den Schwerpunkten Aufnahme und Postproduktion.

Ausgewählte Prozessbetrachtungen
von der Aufnahme bis zur Rezeption

Stereoskope HD-Produktion

ISBN-10: 3-9810384-8-7
ISBN-13: 978-3-9810384-8-4



DIGITALE
MEDIEN



DIGITALE
MEDIEN

Eine Schriftenreihe der Fakultät Digitale Medien an der Hochschule Furtwangen

Nikolaus Hottong, Paul Walter (Hrsg.)

Fakultät Digitale Medien
Arbeitspapier Nr. 7

Alle Rechte, insbesondere das Recht der Vervielfältigung und Verbreitung, sowie der Übersetzung, vorbehalten. Kein Teil des Werkes darf in irgendeiner Form (durch Fotokopie, Mikrofilm oder ein anderes Verfahren) ohne schriftliche Genehmigung der Fakultät reproduziert oder unter Verwendung elektronischer Systeme (inkl. Online-Netzwerken) gespeichert, verarbeitet, vervielfältigt oder verbreitet werden.

Copyright © 2011 Fakultät Digitale Medien | Hochschule Furtwangen

ISBN-10: 3-9810384-8-7
ISBN-13: 978-3-9810384-8-4

REDAKTION: Prof. Nikolaus Hottong, Paul Walter
COVER: Paul Walter
SATZ: Paul Walter
DRUCK: Druckerei Hochschule Furtwangen

Nikolaus Hottong, Paul Walter (Hrsg.)

Stereoskope HD Produktionen

Ausgewählte Prozessbetrachtungen von
der Aufnahme bis zur Rezeption

Schriftenreihe Fakultät Digitale Medien
HFU Furtwangen University
Juli 2011

Inhaltsverzeichnis

Vorwort der Herausgeber	15
NIKOLAUS HOTTONG/PAUL WALTER	

I. Einleitung: Einführung in die zentralen Basisbegriffe der S3D-Produktion	19
NIKOLAUS HOTTONG/ERIC SCHUMACHER	

Abbildungsverzeichnis.....	21
----------------------------	----

1. Zentrale Basisbegriffe der Stereoskopie	23
1.2 Aufnahme stereoskopischer Bilder	28
1.3 Postproduktion stereoskopischer Bilder	30
1.4 Wiedergabe stereoskopischer Bilder	32
1.5 Aktuelle Situation und Einsatzbereiche.....	41

Literaturverzeichnis	44
----------------------------	----

II. Konzeption und Aufbau eines einfachen praxisorientierten S3D-Kamerasystems für Forschungs- und Lehrzwecke	49
DIRK HENSEL/PAUL WALTER	

Abbildungsverzeichnis.....	51
----------------------------	----

1. Einleitung	53
2. Technische und gestalterische Parameter	56
2.1 Stereobasis	56
2.2 Brennweite.....	56
2.3 Scheinfensterweite.....	57

2.4 Kadrierung	57
2.5 Tiefenschärfe	58
2.6 Parallel oder Konvergent	59
2.7 Kamerafunktionen	61
3. Ausgangssituation	62
3.1 Vorhandenes Equipment	62
3.2 Vorausgehende Erkenntnisse	62
3.3 Notwendiges Equipment	63
4. Technische Voraussetzungen	64
4.1 Geeignete Kameras	64
4.2 Objektiv	67
4.3 Kamera- und Objektivsteuerung	71
4.4 Synchronisation	72
4.5 Kameramontage	74
4.6 Datentransfer	76
4.7 Aufnahme	78
4.8 Vorschaumöglichkeiten	82
4.9 Stromversorgung	84
5. Praktische Umsetzung Version 1	87
5.1 Objektiv	87
5.2 Rig	87
5.3 Kameratests	89
5.4 Bewertung der Ergebnisse	104
6. Praktische Umsetzung Version 2	105
6.1 Kaufentscheidung Kamera: AVT GE1910C	105
6.2 Objektiv CVO GMMPZ1664MCN	105
6.3 Aufnahme- Datenübertragung	106
6.4 Stereobasis	107
7. Zusammenfassung	109
Literaturverzeichnis	110

III.	Eine Einführung in die Bildverarbeitung von stereoskopischen Views	113
	ERIC SCHUMACHER	
	Abbildungsverzeichnis.....	115
	Tabellenverzeichnis	117
1.	Einführung	119
1.1	Das digitale Bild.....	119
1.2	Digitale Videoformate.....	120
2.	Kameraparameter	122
2.1	Bildbestimmende Kameraparameter	122
2.2	Zusätzliche Kameraparameter	141
2.3	Kameraparameter eines Stereosystems.....	144
2.4	Zusammenfassung.....	146
3.	Projektion und stereoskopische Projektion	148
3.1	Projektion	148
3.2	Projektion bei einer einzelnen Kamera	149
3.3	Projektion bei einem Stereosystem	158
4.	Zusammenfassung	190
	Literaturverzeichnis	192

IV.	Stereoskopische Fusion von CG-Environments und Real-Footage Elementen.....	197
	FRANK HOFMEYER	
	Abbildungsverzeichnis.....	199
1.	Einleitung	201
1.1	Postproduktion und Visual Effects	201
1.2	Compositing	202
1.3	Chroma-Keying	203
1.4	Rotoscoping	206
2	Stereoskopische Fusion in 2D	209
2.1	Planung und Vorbereitung.....	210
2.2	Realaufnahme	213
2.3	Postproduktion	214
2.4	Stereoskopische Wiedergabe	231
2.5	Praktische Erkenntnisse und Lösungen	233
2.6	Bewertung des 2D-Fusionsverfahrens.....	241
3.	Alternative Fusionsverfahren	242
3.1	Darsteller als Textur: 3D-Fusionsverfahren	242
3.2	Planare Stereoskopie: 2.5D-Fusionsverfahren	249
4.	Diskussion der Ergebnisse	258
4.1	Ausgangssituation.....	258
4.2	Diskussion	260
4.3	Ergebnisse der Diskussion.....	262
4.4	Fazit	263
	Literaturverzeichnis	266

V.	Untersuchungen zur automatischen Generierung von Tiefenkarten in der S3D-Produktion.....	281
	ERIC SCHUMACHER	
	Abbildungsverzeichnis.....	283
	1. Einführung.....	285
	2. Tiefenbilder	289
	2.1 Prinzipien der Tiefenbildaufnahme	289
	2.2 Gewinnung von Tiefeninformation aus Stereobildern.....	293
	2.3 Zusammenfassung.....	314
	3. Tests.....	315
	3.1 Planung.....	315
	3.2 Ablauf.....	317
	3.3 Nachbearbeitung	317
	4. Zusammenfassung und Fazit.....	337
	Literaturverzeichnis	338
VI.	Untersuchungen zur H264 Multi-view Video Coding Ex- tension für den Einsatz in der Distribution von S3D-Filmen	343
	MARTIN MAURER	
	Abbildungsverzeichnis.....	345
	Tabellenverzeichnis	347
	1. Einführung.....	349
	2. Stereskope Speicher- und Übertragungsprinzipien	351

2.1. DCDM	352
2.2. Side-by-Side	352
2.3. Over-and-Under	353
2.4. Field-Sequential/Interlaced-Stereo	354
2.5. Sensio	355
2.6. Schnittstellt HDMI	358
3. Multi-view Video.....	359
3.1. Definition	359
3.2. Historische Entwicklung	359
3.3. Anwendungsszenarien.....	359
3.4. MVC-Architektur	362
3.5. Anforderungen an MVC	363
4. Stereoskopische Videocodierung mit H264/MVC	365
4.1. Allgemein	365
4.2. Videocodierung mit H264/AVC	365
4.3. MVC-Erweiterungen gegenüber H264/AVC	367
5. Der JMVC-Codec in der Praxis	372
5.1. Überblick	372
5.2. Installation und Kompilierung	372
5.3. Bestandteile des JMVC-Softwarepaket	373
5.4. Testdateien	378
5.5. Testsoftware	379
5.6. Testhardware	379
6. Messergebnisse	381
6.1. Anforderungskatalog.....	381
6.2. Durchführung.....	382
6.3. Abwärtskompatibilität	396
6.4. Fazit	397
Literaturverzeichnis	398

VII.	S3D-Grammatik: Analysen, Einordnungen und Definitionsansätze zur Bildsprache des stereoskopischen Films.....	401
	PAUL WALTER	

Abbildungsverzeichnis.....	403
----------------------------	-----

1. Einleitung	405
2. Historische Entwicklungszyklen im Kino.....	407
2.1 Tonfilm	407
2.2 Farbfilm	407
2.3 Breitwandfilm	408
2.4 Entwicklung des 3D-Stereofilms	409

3. Grundkonzepte der Filmwahrnehmung und Filmpsychologie	412
3.1 Filmerlebnis	412
3.2 Dynamik der menschlichen Wahrnehmung	412
3.3 Wahrnehmungsgrenzen	414
3.4 Filmwahrnehmung und Informationsdichte.....	415
3.5 Informationsverarbeitung im Film	415
3.6 Emotional bindende Reizgefühle	416
3.7 Stereofilm-Benefits: Immersion oder Theatralik?	417

4. Bildsprache - Definition und Einordnung.....	419
4.1 Bildsprache - Definition	419
4.2 Denotation und Konnotation	420
4.3 Kinematografische Grammatik und Syntax	422
4.4 Stereo als kinematografischer Code?	423

5. Bildsprache des 3D-Stereofilms	425
5.1 Inszenierung	425
5.2 Spielraum oder Raumspiel	425
5.3 Klassische Organisationsprinzipien für Raum und Zeit	426

5.4 Stereo und Filmdauer: Effekt oder Stilmittel	432
5.5 Rückwirkungen der Stereoskopie auf tradierte kinematographische Codes	432
5.6 Neue Gestaltungs-, Steuerungsmöglichkeiten der Stereografie	434
5.7 Dramaturgische Grundkonzepte des Stereofilms	435
5.8 Analyse ausgewählten Produktionsbeispiele 2009.....	437
 6. Kritiken am 3D-Stereofilm	439
6.1 Stereofilme als Zuschauermagnet	439
6.2 Brillenproblematik	439
6.3 Irritationen beim S3D Seherlebnis.....	440
6.4 Inhaltliche Einsätze von S3D	441
 7. Fazit	442
 Literaturverzeichnis	444
 Autorenbeschreibung	447

Vorwort der Herausgeber

Dies ist der siebte Arbeitsbericht der Fakultät Digitale Medien an der Hochschule Furtwangen. Wie alle vorherigen stellt auch dieser Bericht eine Querschnittsaufnahme unserer praxisbezogenen Medien-Forschung dar.

Nach zwei Jahren haben wir wieder die Gelegenheit ergriffen, aktuelle Projektstände im Forschungsgebiet Stereoskope HD-Produktion zu dokumentieren und in Druckform für interessierte Leser zugänglich zu machen. Diesmal konnten wir einen schönen thematischen Bogen von der S3D¹-Aufnahme (Kamera) über die Postproduktion bis zur Distribution (Multiview-Codec) spannen, abgerundet durch eine Betrachtung zur Bildsprache und Grammatik von Stereofilmen.

Der Arbeitsbericht fasst die neusten Ergebnisse unserer Forschungs- und Entwicklungsarbeiten zusammen, welche sich seit dem Arbeitsbericht Nr. 5 ergeben haben. Studierende der Hochschule Furtwangen haben dafür Teile ihrer erfolgreich abgeschlossenen Thesisarbeiten beigetragen. Diese Teile wurden von uns überarbeitet, zusammengeführt und auf den aktuellen Stand der laufenden Forschungsarbeiten ergänzt. Das Ergebnis halten Sie hiermit in Händen.

Um die Lust auf das Lesen anzufachen möchten wir kurz aufzeigen wie dieser Bericht strukturiert ist:

Die Eröffnung bestreiten Nikolaus Hottong und Eric Schumacher mit einer kurzen Einführung in die technischen Grundlagen der Stereoskopie. (Anm.: Lesern, die sich neu der mit der Thematik befassen legen wir den vorausgegangenen Arbeitsbericht Nr.5 zur Lektüre nahe, welcher sich en detail mit den S3D-Grundlagen auseinandersetzt).

Danach beschreiben Dirk Hensel und Paul Walter die Konzeption und den Aufbau eines einfachen, praxisorientierten S3D-Kamerasystems für Forschungs- und

1 S3D: Kurzform für stereoskopisch produzierte Bilder mit verstärkter 3D-Tiefenwirkung

Lehrezwecke und dokumentieren das im Jahre 2010 an der Fakultät DM in Betrieb genommene S3D-Kamerasystem.

Unsere Arbeiten zur S3D-Postproduktion bilden einen deutlichen Schwerpunkt in diesem Bericht. Eric Schumacher führt daher zunächst grundsätzlich in die Bildverarbeitung von stereoskopischen Views ein.

Frank Hofmeyer beschreibt daraufhin die unterschiedlichen Postpro-Möglichkeiten zur stereoskopischen Fusion von S3D-Realaufnahmen mit CG-Environments und zeigt exemplarisch die Vor- und Nachteile der unterschiedlichen Ansätze auf. Tiefenkarten nehmen heute in der S3D-Postproduktion eine Schlüsselposition zur Kontrolle und Manipulation von Stereobildern ein. Eric Schumacher setzt sich daher im Anschluss mit der Frage auseinander, wie diese „z-maps“ erfolgreich aus einer S3D-Realaufnahme extrahiert werden können und zeigt auf, welche konkreten Probleme damit in der Produktionspraxis verbunden sind.

Nachfolgend berichtet Martin Maurer von seiner Annäherung an die 2009 veröffentlichte Multi-view Video Coding Extension des H.264-Videokompressionsstandards und seine ersten praktischen Untersuchungen zur Codiereffizienz bzw. der resultierenden Bildqualität dieser Extension, welche mittlerweile für die 3D-Bluray Disc zur Norm erhoben wurde.

Paul Walter schließt diesen Bericht mit einem ganz neuen - aber wichtigen - Untersuchungsaspekt ab. Wir produzieren S3D-Filme letztendlich nur für eine Zielgruppe: Menschen. Er beschäftigte sich daher mit Analysen, Einordnungen und Definitionsansätzen zur Bildsprache des stereoskopischen Films und der spannenden Frage, ob für die Kommunikation mittels Stereofilm eine S3D-Grammatik hergeleitet werden kann. Die Antwort dazu finden Sie am Ende dieses Berichtes.

Schlussendlich sind es, alles in allem und zu unserer eigenen Überraschung, über 400 Seiten Informationen zum Thema geworden.

Wir möchten uns daher mit dieser Druckausgabe ganz herzlich bei allen Beteiligten für das Engagement und die Zusammenarbeit bedanken. Die Arbeiten werden natürlich fortgesetzt.

Furtwangen, im Juli 2011
Nikolaus Hottong, Paul Walter

Teil 1

Einleitung: Einführung in die zentralen Basis- begriffe der S3D-Produktion

Nikolaus Hottong, Eric Schumacher

Abbildungsverzeichnis

Abbildung 1: Horopter	27
Abbildung 2: Stereorigs	29
Abbildung 3: Stereoaufnahme	30
Abbildung 4: Parallaxe	34
Abbildung 5: Polfilter	36
Abbildung 6: Infitec	37
Abbildung 7: Parallax-Barriere.....	40
Abbildung 8: Lentikular-Display	40

1 Zentralen Basisbegriffe der Stereoskopie

1.1 Visuelle Wahrnehmung räumlicher Tiefe beim Menschen

Ein Mensch mit einem intakten System zur visuellen Wahrnehmung und Interpretation seiner Umgebung ist durch dieses System dazu befähigt, Aussagen über die räumliche Tiefe besagter Umgebung zu machen. Um räumliche Tiefe über den visuellen Weg wahrzunehmen, nutzt der Mensch, bzw. sein Gehirn eine Vielzahl sog. Hinweisreize oder Tiefenhinweise (depth cues) – in den meisten Fällen mehrere davon gleichzeitig und unbewusst¹. Goldstein teilt diese in drei Kategorien ein:²

- monokulare Tiefenhinweise
- okulomotorische Tiefenhinweise
- stereoskopisches Sehen

Monokulare Tiefenhinweise und das stereoskopische Kriterium basieren direkt auf optischen Informationen. Während die erste Hinweisart auch mit nur einem Auge wahrzunehmen ist, werden für das stereoskopische Sehen zwingend beide Augen benötigt³. Okulomotorische Tiefenhinweise liefern Informationen, die lediglich indirekt auf optischen Veränderungen basieren. Das Gehirn entnimmt die Hinweise aus Muskelveränderungen, die als Reaktion der Augen auf Sichtbares erfolgen⁴.

1 Vgl. Lipton 1982, S. 60f.

2 Vgl. Goldstein 2008, S. 186.

3 Vgl. Lipton 1982, S. 60.

4 Vgl. Lipton 1982, S. 58.

1.1.1 Monokulare Tiefenhinweise

1.1.1.1 Bildbezogene monokulare Tiefenhinweise

- Texturdichtegradient ⁵: Oberflächen mit feinen Details oder Mustern scheinen mit zunehmender Entfernung an Detaillierung zu verlieren. Dieser Übergang von feinen Strukturen zu einheitlich wirkenden Flächen stellt ein Indiz für Tiefe dar.
- Verdeckung ⁶: Objekte, die sich vor anderen Objekten befinden, verdecken diese entweder vollständig oder teilweise. Die Art, wie sich Objekte in einer Szene gegenseitig verdecken, liefert Informationen bezüglich deren räumlicher Staffelung.
- Atmosphärische Perspektive ⁷: Objekte, die über weite Ferne hinweg erkennbar sind wie beispielsweise Bergketten, wirken kontrastlos und detailarm. Dies liegt an den Partikeln, die sich zwischen dem Betrachter und dem fixierten Objekt in der Luft befinden. Gasmoleküle, Staub, Rauch usw. reflektieren oder absorbieren einen Teil der vom Objekt reflektierten Lichtstrahlen und reduzieren dadurch Kontraste und Details. Besonders gut lässt sich dieser Effekt beobachten, wenn Aussichten bei klarem Wetter mit Aussichten bei nebligem oder trübem Wetter verglichen werden.
- Beleuchtung und Schattenwurf ⁸: Die Art, wie ein Objekt bei einer bestimmten Beleuchtungssituation Lichtstrahlen reflektiert und Schatten wirft, gibt Hinweise auf seine räumliche Gestalt und Oberflächenbeschaffenheit.
- Relative Höhe im Blickfeld ⁹: Wenn ein Objekt innerhalb des Blickfeldes eines Betrachters sehr nahe am Horizont liegt, scheint es weit entfernt. Objekte, die näher am oberen oder unteren Rand des Sichtfeldes liegen, scheinen sich näher beim Betrachter zu befinden.
- Relative Größe im Blickfeld ¹⁰: Objekte gleicher Größe nehmen abhängig von ihrer Entfernung zum Betrachter mehr oder weniger Raum in dessen Blickfeld ein. Objekten, welche mehr Raum einnehmen, werden kurze Distanzen zugeschrieben und Objekte mit kleinerem Raumanspruch werden als weiter entfernt interpretiert.
- Bekannte Größe ¹¹: Die Wahrnehmung von räumlicher Tiefe hängt zusammen mit der persönlichen Erfahrung. Ist die Größe eines bestimmten Objekts

5 Vgl. Lipton 1982, S. 56.

6 Vgl. Goldstein 2008, S. 187.

7 Vgl. Mahler 2005, S. 277.

8 Vgl. Schmidt 2009, S. 27.

9 Vgl. Goldstein 2008, S. 187.

10 Vgl. Schmidt 2009, S. 27.

11 Vgl. Goldstein 2008, S. 188.

bekannt, so kann ein Betrachter auf Grund dieses Wissens Aussagen über die Tiefenposition des Objekts machen.

- Perspektivische Konvergenz ¹²: Parallel verlaufende Linien scheinen mit zunehmender Entfernung aufeinander zuzulaufen und sich in einem Schnittpunkt zu treffen. Diese perzeptuelle Annäherung hilft dem Menschen ebenfalls dabei, die räumliche Struktur einer Szene zu bestimmen.

1.1.1.2 Bewegungsinduzierte monokulare Tiefenhinweise

- Bewegungsparallaxe ¹³: Dieser Effekt tritt auf, wenn sich ein Beobachter seitlich parallel zu einem Objekt bewegt und er den Blick dabei auf dieses Objekt gerichtet hält. Als konkretes Beispiel ist hier der Blick aus dem Seitenfenster eines fahrenden Autos oder Zuges zu nennen. Es scheint, als ob sich nahe Objekte schneller vorbeibewegen als weiter entfernte Objekte. Die Bewegungsgeschwindigkeit einzelner Objekte liefert dadurch Hinweise zu deren Entfernung. Eine ähnliche Beobachtung entsteht in einer Situation, in der sich eines oder mehrere Objekte parallel zum Betrachter bewegen, dieser jedoch seine Position nicht verändert.

- Dynamische Verdeckung ¹⁴: Wenn in einer Szene Bewegungen erfolgen, kann die Situation entstehen, dass sich die sichtbaren Objekte zeitweise gegenseitig vollständig oder teilweise verdecken. Je nach Art der Bewegungen und der Position der Objekte erfolgen Verdeckungen nur über einen bestimmten Zeitraum hinweg oder die Verdeckungsverhältnisse ändern sich über den Zeitverlauf. Durch das Beobachten dieser Vorgänge ergeben sich für den Betrachter Hinweise auf die räumliche Tiefe der Szenenobjekte.

1.1.2 Okulomotorische Tiefenhinweise

- Akkommodation ¹⁵: Die Linse im menschlichen Auge besitzt keine starre Form, über Muskelkontraktionen kann ihre Form verändert werden. Dieser Vorgang, der einfallendes Licht gezielt auf die Fovea lenkt, wird durch die veränderbare Gestalt der Linse ermöglicht. Wird auf ein Objekt fokussiert, welches sich nahe beim Betrachter befindet, muss die Linse dicker und kugelförmiger werden, um die Lichtstrahlen so zu brechen, damit sie in die Fovea gelangen. Bei einer größeren Objektentfernung fallen die Lichtstrahlen unter einem anderen Winkel in das Auge, die Linse muss daher länger und schmäler werden, um die Lichtstrahlen korrekt brechen zu können. Aus dem Kontraktionszustand der beteiligten Muskelgruppen entnimmt das Gehirn Informationen bezüglich

¹² Vgl. Schmidt 2009, S. 27.

¹³ Vgl. Lipton 1982, S. 56.

¹⁴ Vgl. Goldstein 2008, S. 190.

¹⁵ Vgl. Lipton 1982, S. 58.

des Zustandes der Linse und auf welche ungefähre Entfernung diese folglich ausgerichtet sein muss.

- Konvergenz ¹⁶: Hier entnimmt das Gehirn Informationen aus der muskulären Drehung der Augen. Wird ein Objekt fixiert, welches sich in kurzer Distanz befindet, drehen sich die Augen nach innen in Richtung der Nasenspitze. Auch dieser Vorgang dient dazu, die Lichtstrahlen des Objekts auf die Fovea zu leiten. Bei einer größeren Objektdistanz (ca. 60 m) erfolgt keine Augendrehung nach innen, die Sehachsen sind nahezu parallel ausgerichtet ¹⁷. Die Drehung erfolgt ebenfalls über Muskelgruppen, die dem Gehirn über ihren Kontraktionszustand Hinweise zur Ausrichtung der Augen und somit Informationen in Bezug auf die Distanz des fixierten Objekts liefern.

1.1.3 Stereoskopisches Sehen

Zwei Augen betrachten eine Szene aus leicht unterschiedlichen Perspektiven. Dies lässt sich leicht anhand eines Experiments nachvollziehen, bei dem ein Zeigefinger auf Höhe der Augen und in einer Entfernung von ca. 30 cm gehalten wird. Wird der Zeigefinger abwechselnd nur mit dem linken oder rechten Auge betrachtet, so ist ein „Hin-und-her-Springen“ des Fingers erkennbar. Daraus folgt, dass auf den beiden Netzhäuten Bilder entstehen, die auf Grund des perspektivischen Unterschiedes nicht exakt identisch sind, sondern Differenzen aufweisen ¹⁸. Einzelne Objektpunkte der Szene werden also auf zwei Bildpunkten auf den Netzhäuten abgebildet. Wenn ein Betrachter seinen Blick auf einen Objektpunkt F fixiert, so befindet sich dieser Punkt in einer bestimmten Entfernung. Solange F fixiert wird, werden seine beiden zugehörigen Bildpunkte f und f' auf den Netzhäuten an korrespondierenden Stellen abgebildet ¹⁹, d.h. die Position der Punkte wäre identisch, wenn die Netzhäute übereinander liegen würden. Der sog. optische Horopter ist ein gedachter Kreis, der durch die Mittelpunkte der beiden Augenoptiken und den Fixationspunkt F verläuft. Sämtliche Objektpunkte, die sich außer F zusätzlich auf dem Horopter befinden, werden ebenfalls als korrespondierende Netzhautpunkte abgebildet. Die Gestalt des Horopters ändert sich natürlich, sobald ein neuer Objektpunkt fixiert wird, – also fast ständig – das Prinzip bleibt jedoch erhalten. Objektpunkte im Blickfeld, die sich in einer kleineren oder größeren Distanz befinden als der aktuell fixierte Punkt, liegen nicht auf dem Horopter. Sie erzeugen auf den Retinae keine korrespondierenden Bildpunkte, ihre Position ist auf der linken Netzhaut eine andere als auf der rechten. Dieser Versatz wird als Disparität δ , genauer als binokulare Disparität oder Querdisparität bezeichnet ²⁰. Bei Objektpunkten, die sich in einer weiteren Entfernung befinden als der Verlauf des Horopters, entsteht ungekreuzte Querdisparität, d.h. die Bildpunkte

¹⁶ Vgl. Mahler 2005, S. 277.

¹⁷ Vgl. Röder 2007, S. 6.

¹⁸ Vgl. Goldstein 2008, S. 192.

¹⁹ Vgl. Lipton 1982, S. 105f.

²⁰ Vgl. Mahler 2005, S. 277f.

liegen weiter innen auf den Netzhäuten. Objektpunkte, die vor dem Horopter liegen, erzeugen gekreuzte Querdisparität; die Bildpunkte liegen dabei weiter außen auf den Retinae. Diese versetzte Lage der Bildpunkte auf den Netzhäuten liefert dem Gehirn über die Querdisparitätsverhältnisse Informationen hinsichtlich der räumlichen Tiefe dessen, was sich im Blickfeld befindet ²¹. Der Faktor der Disparität spielt ebenfalls eine Rolle bei der technischen Reproduktion des stereoskopischen Sehens.

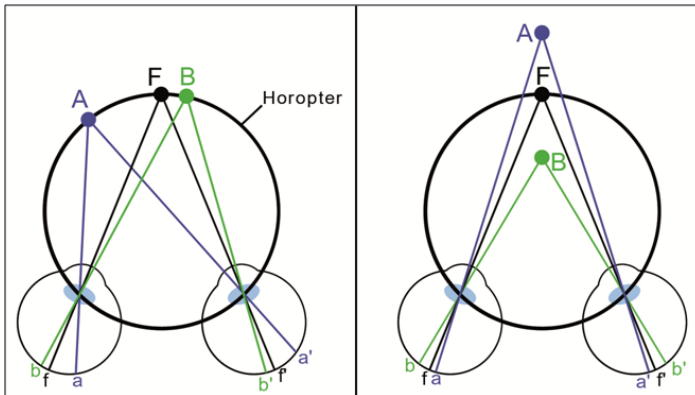


Abbildung 1: Horopter

1.1.4 Technische Stereoskopie

Bei der technischen Stereoskopie wird versucht, das binokulare Sehen mit technischen Mitteln zu reproduzieren. Hierzu werden mit entsprechenden Bildaufzeichnungsgeräten von derselben Szene zwei Aufnahmen aus zwei Perspektiven gemacht ²². Die Aufnahme beider Bilder muss zeitgleich erfolgen, i.d.R. werden also zwei Kameras benötigt. Eine Ausnahme besteht bei Szenen, in denen keinerlei Veränderung erfolgt. In diesem Fall können beide Bilder aus verschiedenen Perspektiven nacheinander mit nur einer Kamera aufgenommen werden. Es können sowohl Standbilder (Fotografie) als auch Bewegtbild-Sequenzen (Film, Video) stereoskopisch aufgenommen werden. Die beiden Aufnahmen repräsentieren die unterschiedlichen Perspektiven der Augen. Die spätere Wiedergabe der Inhalte muss daher so erfolgen, dass jedem Auge des Betrachters ausschließlich die entsprechende Ansicht präsentiert wird ²³. Es darf nicht passieren, dass dem linken Auge die rechte Ansicht geliefert wird oder umgekehrt. Wird die Wiedergabe korrekt durchgeführt, so kann das Gehirn des Betrachters die beiden Ansichten zu einem Gesamtraumbild fusionieren und das Gesehene erscheint plastisch und mit einer wirklichen räumlichen Tiefe. Es existieren darüber hinaus Auf-

²¹ Vgl. Goldstein 2008, S. 194.

²² Vgl. Röder 2007, S. 7.

²³ Vgl. Röder 2007, S. 22.

nahme- und Wiedergabeverfahren, bei denen mehr als zwei Ansichten aufgenommen werden. In diesem Arbeitspapier werden lediglich Verfahren mit zwei Ansichten behandelt. Geht der Inhalt darüber hinaus, so wird dies explizit erwähnt.

1.2 Aufnahme stereoskopischer Bilder

Wenn die Aufnahme beider Ansichten zeitgleich erfolgen soll, werden in den meisten Fällen spezielle Stereo-Rigs benötigt. Es handelt sich dabei um Konstruktionen, die den gleichzeitigen Einsatz zweier Kameras ermöglichen. Hinsichtlich der Ausrichtung der Kameras finden in der Praxis hauptsächlich zwei Varianten Verwendung. Bei der ersten Variante verlaufen die Strahlengänge parallel zueinander. Hierfür werden sog. Side-by-Side-Rigs oder Mirror-Rigs eingesetzt²⁴. Bei Side-by-Side-Rigs werden die Kameras horizontal nebeneinander platziert, z.B. auf einer sog. Stereoschiene. Bei Mirror-Rigs sind die Kameras so befestigt, dass ihre optischen Achsen zunächst in einem rechten Winkel zueinander verlaufen. Ein halbdurchlässiger Strahlenteiler sorgt dafür, dass das Licht der Szene korrekt in die Kameraobjektive gelenkt wird. Der Grund für diesen Konstruktionsaufwand hängt mit dem Parameter der sog. Stereobasis SB zusammen. Dieser Wert beschreibt die Distanz zwischen den optischen Achsen beider Kameras, genauer gesagt die Distanz zwischen den beiden optischen Zentren²⁵. In vielen Fällen wird eine Stereobasis von ca. 62 – 65 mm gestrebt, da dies dem durchschnittlichen horizontalen Augenabstand entspricht²⁶. Aufgrund der physikalischen Größe der Objektive ist dies bei Side-by-Side-Rigs oft nicht zu realisieren. Mirror-Rigs erlauben jedoch bedingt durch ihren Aufbau kleinere Basisabstände, sogar im Bereich weniger Millimeter.

Bei der zweiten Variante der Kameraanordnung werden die Kameras ähnlich der Side-by-Side-Anordnung horizontal nebeneinander platziert. Darüber hinaus werden diese jedoch zueinander gedreht, so dass die optischen Achsen nicht parallel verlaufen, sondern konvergent und sich in einem bestimmten Punkt schneiden²⁷. In manchen Fällen wird durch eine solche konvergente Ausrichtung eine bessere Betrachtungsqualität erreicht als bei parallel ausgerichteten Kameras²⁸. Problematisch wird diese Variante, wenn sich die Augen auf Punkte konzentrieren, welche hinter dem Schnittpunkt der optischen Achsen liegen. Durch die Ausrichtung wird in diesen Bereichen der perspektivische Unterschied zwischen den Ansichten so groß, dass das Gehirn die beiden Ansichten nicht mehr fusionieren kann oder dass die Augen gezwungen werden, sich nach außen in Richtung der Schläfen zu drehen (Divergenz)²⁹. Dies erzeugt sehr schnell Unwohlsein beim Betrachter.

²⁴ Vgl. Maier Nov. 2008.

²⁵ Vgl. Schreer 2005, S. 67.

²⁶ Vgl. Schmidt 2009, S. 448.

²⁷ Vgl. Lipton 1982, S. 95.

²⁸ Vgl. Zone 2005, S. 21.

²⁹ Vgl. Lipton 1982, S. 102.

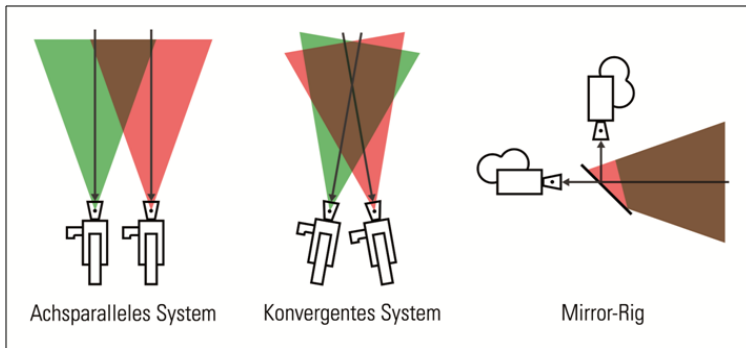


Abbildung 2: Stereoreizung

Beim Einsatz von zwei Kameras müssen verschiedene Bedingungen erfüllt sein:

- Die Kameras müssen vom gleichen Modell sein, da unterschiedliche Modelle konstruktionsbedingt durch unterschiedliche Abbildungsverhältnisse zu unterschiedlichen Stereoansichten führen.
- Beide Kameras müssen identische intrinsische und extrinsische Parametereinstellungen aufweisen.
- Bei beiden Kameras muss sich die Aufnahme synchron starten und stoppen lassen, bzw. es muss gewährleistet sein, dass die Synchronität zwischen beiden Ansichten hergestellt werden kann.
- Alle Parameteränderungen während der Aufnahme (z.B. Zoom) müssen ebenfalls absolut synchron erfolgen.

Die Stereobasis wird zusammen mit der verwendeten Brennweite zur Berechnung der sog. Scheinfensterweite (kurz: SFW) herangezogen. Das Scheinfenster hat seinen Namen von der Tatsache, dass man beim Betrachten eines Stereobildes den Eindruck hat, durch ein imaginäres Fenster zu blicken. Das Scheinfenster teilt das Stereobild in Mittel-, Hinter- ($SFW <$) und Vordergrund ($SFW >$). Während das Scheinfenster die hintere Grenze des Vordergrundbereiches darstellt, so definiert der sog. Nahpunkt die vordere Grenze dieses Bereichs. Es handelt sich dabei um den Objektpunkt der Szene, der die kürzeste Distanz (Nahpunktweite, kurz: NPW) zu den Kameras aufweist³⁰. Ebenso existiert ein Fernpunkt mit einer Fernpunktweite (kurz: FPW), welche die Distanz des am weitesten entfernten Szenenpunktes darstellt. Abhängig von der Scheinfensterweite existiert also bei jeder Stereoaufnahme ein Bereich, in

30 Vgl. Röder 2007, S. 14.

welchem sich Szenenobjekte befinden müssen, um bei der späteren Wiedergabe dreidimensional zu wirken. Befinden sich Objekte in einer Distanz zu den Kameras, welche die Nahpunktweite unterschreitet, so gilt für diese Objekte, dass ihre Abbildungen auf den Stereobildpaaren zu große Disparitäten aufweisen, um vom Gehirn zu einem räumlichen Gesamtbild fusioniert werden zu können. Stattdessen nimmt der Betrachter gleichzeitig zwei flache Einzelbilder wahr, wodurch die Illusion des Stereoeffekts zunichte gemacht wird³¹. Dadurch wird deutlich, dass eine korrekte Stereoaufnahme von einer Vielzahl von Parametern abhängt.

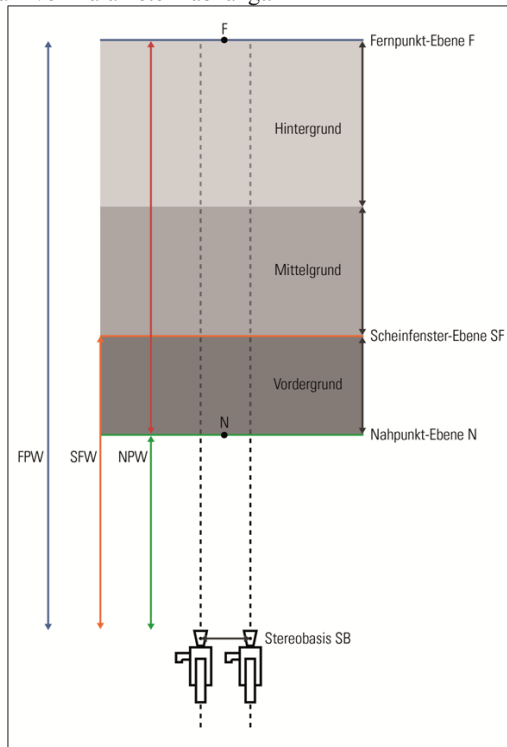


Abbildung 3: Stereoaufnahme

1.3 Postproduktion stereoskopischer Bilder

Wie auch bei der Aufnahme bedeutet eine stereoskopische Produktion für die Nachbearbeitung einen zusätzlichen Aufwand im Vergleich zu Produktionen, bei denen nur eine Ansicht aufgenommen und wiedergegeben wird. Die Besonderheiten dieser

³¹ Vgl. Röder 2007, S. 14.

zusätzlichen Anforderungen lassen sich in zwei Bereiche einteilen. Der erste Bereich umfasst einerseits Prozesse, die dazu dienen, technisch korrekte Stereobildpaare zu erzeugen, um damit Unwohlsein beim Zuschauer zu vermeiden. Zum Zweiten zählen zu diesem Bereich Arbeitsschritte, welche eine Nachbearbeitung hinsichtlich ästhetischer Gesichtspunkte darstellen, um durch das visuelle Empfinden die Dramaturgie ³² zu unterstützen. Zum zweiten Bereich der zusätzlichen Anforderungen zählt die technische Infrastruktur, welche die Arbeitsschritte des ersten Bereichs überhaupt erst ermöglicht ³³.

1.3.1 Gestalterisch-visueller Hintergrund

Bei Stereoproduktionen kommt es oft vor, dass die beiden Ansichten hinsichtlich unterschiedlichster Anforderungen aufeinander abgestimmt werden müssen. Kommt beispielsweise bei der Aufnahme ein Strahlenteiler zum Einsatz so kann bereits dies zu unterschiedlichen Helligkeits- und/oder Farbverteilungen zwischen beiden Ansichten führen, welche nachträglich korrigiert werden müssen ³⁴. Ein anderer Arbeitsschritt besteht in der Korrektur von vertikalen und horizontalen Verschiebungen ³⁵. Auch wenn bei der Aufnahme mit großer Sorgfalt gearbeitet wird, kann es durchaus vorkommen, dass zwischen den beiden Stereoansichten Unterschiede bestehen, z.B. wenn die sich die beiden Kameras aufgrund einer leichten Bodenunebenheit nicht exakt auf gleicher Höhe befanden. In diesem Fall müssen die Ansichten bzw. die Bildausschnitte aneinander angeglichen werden. Weitere konkrete Störfaktoren innerhalb der Stereoaufnahme können durch (einseitig auftretende) Blendenflecke, optische Verzerrungen und Spiegelungen verursacht werden ³⁶. Besondere Sorgfalt ist des Weiteren auch beim Compositing notwendig, d.h. bei der Kombination von mehreren Bildelementen oder der Integration von Objekten, welche mit einem 3D-Computerprogramm erzeugt wurden. Sämtliche Elemente müssen bezüglich ihrer Positionierung, Größe, Beleuchtung, Perspektive und Raumtiefe korrekte Eigenschaften aufweisen, um glaubhaft in das aufgenommene Originalmaterial integriert werden zu können ³⁷. Aufgrund der Tatsache, dass die Ansichten unterschiedliche Perspektiven aufweisen, genügt es in den meisten Fällen nicht, einen Effekt oder ein Element der einen Ansicht zu kopieren und in die zweite Ansicht einzufügen. In der Regel müssen die Anpassungen zweimal vorgenommen werden.

1.3.2 Technischer Hintergrund

Wie eingangs bereits dargelegt, wird eine technische Infrastruktur benötigt, um die im

32 Vgl. Hottong (Hg.) 2009, S. 87.

33 Vgl. Zone 2005, S. 138.

34 Vgl. Seymour 2008.

35 Vgl. Zone 2005, S. 148.

36 Vgl. Daro 2009.

37 Vgl. Seymour 2008.

vorherigen Abschnitt beschriebenen Arbeitsprozesse durchführen zu können. Einer der dabei relevanten Punkte besteht in der Tatsache, dass bei einer stereoskopischen Produktion die doppelte Menge an Speicherplatz benötigt wird, verglichen mit einer klassischen 2D-Produktion, da für jede Szene je zwei Bildsequenzen existieren. Dies wirkt sich natürlich auch auf das Verwalten der Sequenzen aus, da auch für das Überspielen oder Verschieben von Dateien die doppelte Dauer notwendig ist³⁸. Weiterhin werden Bearbeitungsprogramme benötigt, welche dazu in der Lage sind, einen stereoskopischen Workflow zu unterstützen. Dies betrifft beispielsweise den Prozess des Color-Gradings. Hier muss gewährleistet sein, dass Änderungen hinsichtlich der Farben und Kontraste für beide Ansichten übernommen werden. Eine der wichtigsten Funktionen besteht in der Unterstützung einer stereoskopischen Vorschaufunktion mit geringer Verzögerungszeit, im optimalen Fall in Echtzeit. Alternativ dazu müssten für jede vorgenommene Änderung zeitraubende Testscreenings erstellt werden, um das Bearbeitungsergebnis bewerten zu können. Ist der eigentliche Nachbearbeitungsprozess vollendet, so muss für den Export der finalen Sequenzen stets die doppelte Renderzeit berücksichtigt werden, da jedesmal zwei unterschiedliche Ansichten erzeugt werden müssen.

Derzeit existieren einige Lösungen verschiedener Hersteller, die in unterschiedlichen Ausmaßen eine stereoskopische Nachbearbeitung und Ausgabe ermöglichen. Die Produkte sind u.a. in Form von Plug-Ins und Upgrades für bereits bestehende Systeme verfügbar. Zu nennen sind hier iQ und Pablo der Firma Quantel³⁹, Scratch der Firma Assimilate⁴⁰, SpeedGrade der Firma IRIDAS⁴¹ sowie Nuke ab Version 5.0 und Ocula 1 und 2 der Firma The Foundry⁴². Es werden auch komplette Workstations mit spezieller Software angeboten, wie beispielsweise Sid, ebenfalls entwickelt von Quantel⁴³. Viele Postproduktionsfirmen verfügen mittlerweile über mehr oder weniger standardisierte Workflows, bei denen in vielen Fällen auch firmeneigene Anwendungen zum Einsatz kommen. Simon Robinson, der ein Gründungsmitglied von The Foundry ist und dort als leitender Entwickler arbeitet, vertritt allerdings die Auffassung, dass ein stereoskopischer Workflow so flexibel wie möglich gestaltet werden muss⁴⁴. Dies gilt insbesondere dann, wenn viele verschiedene Firmen zusammenarbeiten und einzelne Arbeitsschritte in unterschiedlich organisierten Workflows durchgeführt werden.

1.4 Wiedergabe stereoskopischer Bilder

1.4.1 Bildschirmparallaxe

38 Vgl. Seymour 2008.

39 Vgl. Quantel 2008.

40 Vgl. Assimilate 2009.

41 Vgl. IRIDAS.

42 Vgl. The Foundry (Hg.) 2010.

43 Vgl. Quantel 2008.

44 Vgl. Seymour 2010.

Wenn die beiden Stereohalbbilder auf einer ebenen Fläche wiedergegeben werden, beispielsweise auf einem elektronischen Display oder einer Kinoleinwand, wird ein aufgenommenes Objekt mit einer Ausnahme (keine Parallaxe, siehe unten) nicht deckungsgleich an derselben Stelle auf der Darstellungsebene abgebildet. Die beiden Bilder des Objekts weisen einen horizontalen Versatz auf, der davon abhängig ist, wo sich die Blickachsen des Betrachters im Raum schneiden. Lipton unterscheidet dabei vier Fälle ⁴⁵.

- Negative Parallaxe: Um das Objekt deckungsgleich wahrzunehmen, müssen sich die Augen je nach Wert des Versatzes mehr oder weniger stark nach innen Richtung Nasenspitze drehen. Je weiter diese Drehung ist, desto weiter scheint sich das fixierte Objekt vor der Darstellungsebene zu befinden ⁴⁶. Es handelt sich dann um den bekannten Stereoeffekt, bei dem Gegenstände aus der Leinwand herauszuragen scheinen oder auf den Zuschauer zufliegen.
- Positive Parallaxe: Hier verhält es sich ähnlich wie beim natürlichen Betrachten weit entfernter Objekte; die Sehachsen weisen eine parallele Ausrichtung auf. Das fixierte Objekt wird bei dieser Konstellation als deckungsgleich wahrgenommen, weil sich die korrespondierenden Punkte im Unendlichen zu befinden scheinen ⁴⁷.
- Keine Parallaxe: Der horizontale Versatz zwischen den beiden Abbildungen des fixierten Objekts hat den Wert Null. Die Blickachsen der beiden Augen schneiden sich auf der Höhe der Leinwand oder des Bildschirms und genau dort wird auch das fixierte Objekt wahrgenommen ⁴⁸.
- Divergenz: In diesem Fall geht die Ausrichtung der Blickachsen über eine parallele Stellung hinaus. Die Augen müssen sich in Richtung der Schläfen drehen, um dem Gehirn zu ermöglichen, die beiden Stereohalbbilder deckungsgleich wahrzunehmen. Dieser Vorgang ist für den Betrachter sehr unangenehm und führt schnell zu Unbehagen und Ermüdung der Augen ⁴⁹.

45 Vgl. Lipton 1982, S. 92ff.

46 Vgl. Maier Juli/Aug. 2008, S. 18.

47 Vgl. Lipton 1982, S. 92ff.

48 Vgl. Maier Juli/Aug. 2008, S. 18.

49 Vgl. Lipton 1982, S. 92ff.

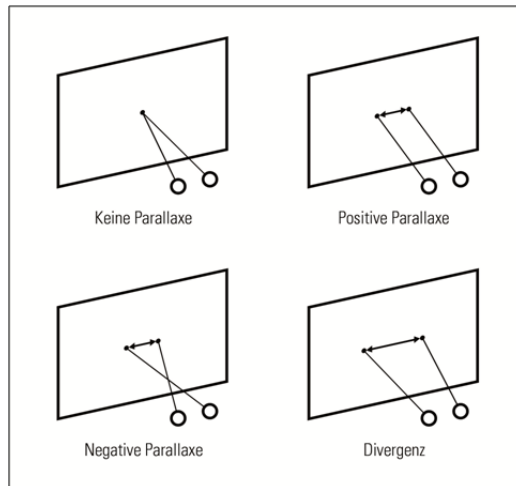


Abbildung 4: Parallaxe

Es wird deutlich, dass bei jeder Stereoproduktion darauf geachtet werden muss, auf welche Weise die spätere Wiedergabe erfolgen wird. Da die Entfernung des Betrachters zur Darstellungsebene und deren Größe Einfluss auf die Tiefenwirkung der Bilder haben, wirken Inhalte unterschiedlich, wenn sie beispielsweise auf einer Kinoleinwand oder einem Fernsehbildschirm betrachtet werden.

Wie zuvor schon ausgeführt, darf bei der Wiedergabe stereoskopischer Inhalte ein bestimmter Betrachtungsabstand nicht unterschritten werden, um die dreidimensionale Wirkung bestmöglich zu erleben. Dieser Minimalabstand ergibt sich im Wesentlichen aus dem Divergenzverbot für die Augenachsen der Betrachtenden. Mittels der vorgestellten Formel lässt sich der Betrachtungsabstand im Verhältnis zu den Einwirkenden Faktoren wie unter anderem Brennweite, Stereobasis, Augenabstand und Objektentfernung berechnen. Die Formel zur Berechnung der Leinwand-Parallaxe bei der Aufnahme wäre laut Lipton:

$$P = (VF \cdot BW \cdot SB) / (NP \cdot FP)$$

Dabei ist VF der Vergrößerungsfaktor, BW die Brennweite, SB die Stereobasis, NP der Nahpunkt und FP der Fernpunkt der Szene. Während der Aufnahme kann man also im Umkehrschluss z.B. die Stereobasis oder die Brennweite berechnen.

Bei der Nachbearbeitung kann die Parallaxe des aufgezeichneten Materials noch einmal korrigiert werden, so dass keine parallaktische Abweichung im Bild über diesem Wert liegt. Dies gilt vor allem für die weit entfernten Objekte im Raum, bei denen die Verschiebung nach der Montage am größten ist.

1.4.2 Wiedergabeverfahren

1.4.2.1 Mit Sehhilfe – passiv

- **Farbfilter (Anaglyphen):** Dieses Verfahren ist eine der ältesten und bekanntesten Methoden zur Wiedergabe stereoskopischer Bildpaare. Dabei kommen Farbfilter zum Einsatz, welche die Eigenschaften von Komplementärfarben ausnutzen. Komplementäre Farben vermischen sich nach den Gesetzen der subtraktiven Farbmischung zu Schwarz⁵⁰; beispielsweise erscheint eine rote Fläche schwarz, wenn sie durch ein cyan-farbenes Glas betrachtet wird und umgekehrt. Dies gilt ebenso für die Kombinationen Blau/Gelb sowie Grün/Magenta. Zur Betrachtung von Stereobildpaaren werden linkes und rechtes Bild mit den jeweils entsprechenden komplementären Farben eingefärbt. Die zugehörige Brille ist mit Farbfilterfolien versehen, die für jedes Auge die „falsche“ Ansicht sperrt⁵¹. Größter Nachteil dieser Methode ist der starke Farbverlust gegenüber der Originalaufnahme. Vorteilhaft ist, dass diese Methode nicht nur bei der Wiedergabe über Displays oder Projektoren funktioniert, sondern auch bei Stereobildern, die auf geeignete Träger gedruckt werden, also beispielsweise Papier oder Plakate.

- **Polarisationsfilter:** Bei Lichtstrahlen handelt es sich bei um elektromagnetische Felder, die sich im Raum ausbreiten und dabei wellenähnliche Schwingungen aufweisen. Diese Schwingungen erfolgen orthogonal zur Ausbreitungsrichtung, wobei sich die Ausrichtung der Schwingungsebene laufend ändert⁵². In diesem Fall handelt es sich um sog. unpolarisiertes Licht. Polarisationsfilter werden eingesetzt, wenn die Ausrichtung der Schwingungsebene einer bewussten Kontrolle unterliegen soll. Wird Licht z.B. linear polarisiert, so werden bis auf eine Schwingungsebene sämtliche anderen unterdrückt. Bei der Wiedergabe von stereoskopischen Inhalten werden zwei Projektoren eingesetzt, von denen einer die linke Ansicht darstellt und der andere die rechte Ansicht. Beide Geräte sind mit je einem Polarisationsfilter ausgestattet, die einen Winkelunterschied von 90° zueinander aufweisen. Die Wiedergabe erfolgt über eine Rückprojektion oder mittels einer speziell reflektierenden Leinwand, welche die Polarisation aufrechterhält⁵³. Der Betrachter nutzt eine Brille mit Polarisationsfiltern, die entsprechend der jeweiligen Ansicht ausgerichtet sind. Auf diese Weise wird das Licht der linken Ansicht vom Filter im rechten Brillenglas herausgefiltert und umgekehrt – jedem Auge wird nur die korrekte Ansicht geliefert⁵⁴. Nachteile dieses Verfahrens sind die eingeschränkte Bewegungsfreiheit des Betrachterkopfes, der vergleichsweise hohe Lichtverlust sowie ggf. die

50 Vgl. Böhringer; Bühler; Schlaich 2008, Bd. 2, S. 207.

51 Vgl. Mahler 2005, S. 280.

52 Vgl. Schröder; Treiber 2007, S. 37.

53 Vgl. Röder 2007, S. 26.

54 Vgl. Mahler 2005, S. 281.

teure Spezialleinwand. Der Vorteil ist die naturgetreue Wiedergabe von Farben und die Möglichkeit, vielen Zuschauern gleichzeitig die Stereoansicht zu liefern. Zur Optimierung des Kopfbeweglichkeit werden heute meist zirkulare polarisationsverfahren angewendet.

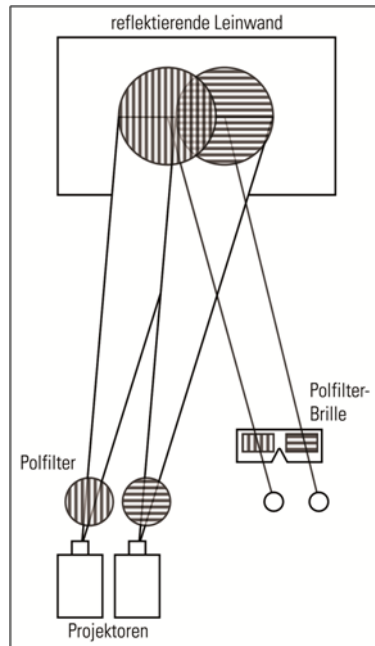


Abbildung 5: Polfilter

- Pulfrich-Verfahren: Dieses Verfahren wurde nach dem deutschen Physiker Carl Pulfrich benannt und nutzt die Tatsache, dass dunkle optische Reize vom Gehirn etwas langsamer wahrgenommen werden als helle ⁵⁵. In der Praxis wird eine Brille verwendet, bei der eines der beiden Gläser stark abgedunkelt ist. Wird dann mit beiden Augen z.B. eine Filmsequenz betrachtet, in der sich ein Gegenstand zur Seite bewegt, bzw. in der die Kamera während der Aufnahme eine entsprechende Bewegung (Parallelfahrt) vollzogen hat, erreicht das Bild, welches das Auge durch das abgedunkelte Glas betrachtet, das Gehirn etwas später als das andere Bild. Durch diesen minimalen zeitlichen Versatz nimmt das Gehirn den Gegenstand an zwei leicht auseinanderliegenden Orten wahr und interpretiert anhand dieser scheinbaren Bewegungsparallaxe eine Tiefe in

⁵⁵ Vgl. Pulfrich 1922.

das Gesehene⁵⁶. Dieses Verfahren liefert daher lediglich einen pseudo-Stereo-effekt. Zudem funktioniert es lediglich dann, wenn sich Bildinhalte bewegen bzw. die Kamera bewegt wird. Darüber hinaus geht durch das abgedunkelte Glas ein großer Teil der Helligkeit verloren. Ein Vorteil besteht darin, dass bei der Aufnahme lediglich eine Kamera bzw. bei der Wiedergabe nur ein Projektor oder ein Display eingesetzt werden muss⁵⁷.

- Wellenlängenmultiplex (Infitec): Dieses Verfahren wurde ursprünglich von der DaimlerChrysler AG für den Einsatz in Virtual-Reality-Systemen entwickelt⁵⁸. Hier kommen Interferenzfilter zum Einsatz, die sich dadurch auszeichnen, dass sie – ähnlich einem Bandpass⁵⁹ in der Nachrichtentechnik – nur Lichtstrahlen mit einer bestimmten Wellenlänge passieren lassen und alle übrigen längeren oder kürzeren Wellenlängen herausfiltern⁶⁰. Linke und rechte Ansicht werden mit zwei Projektoren stereoskopisch projiziert, indem ihre Grundfarben mit unterschiedlichen Wellenlängen dargestellt werden.

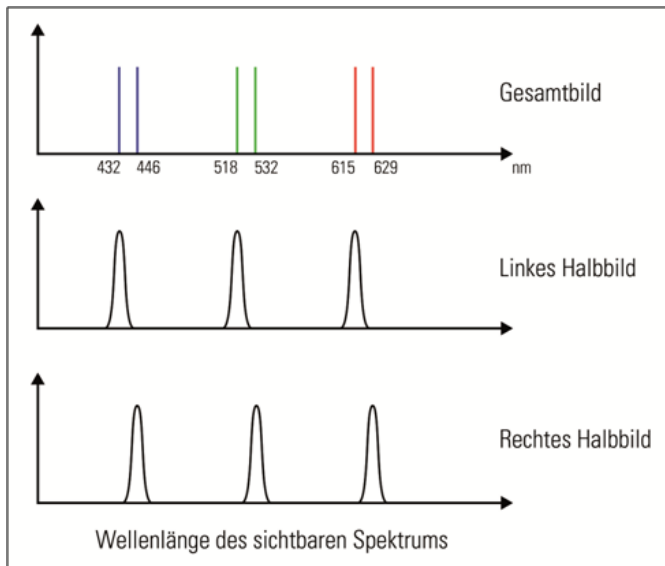


Abbildung 6: Infitec

56 Vgl. Mahler 2005, S. 289.

57 Vgl. Schmidt 2009, S. 27f.

58 Vgl. Wikipedia: Infitec.

59 Vgl. Ohm; Lüke 2005, S. 147.

60 Vgl. Schröder; Treiber 2007, S. 179.

Zur Trennung der gleichzeitig wiedergegebenen unterschiedlichen Bildinhalte dienen Brillen, die mit den entsprechenden Interferenzfiltern ausgestattet sind. Laut Hersteller handelt es sich dabei um eine mehrlagige dielektrische Beschichtung, die auf einen geeigneten durchsichtigen Träger aufgebracht ist, z.B. Quarzglas⁶¹. Beim Einsatz hochwertiger Filter, die sehr trennscharf sind, kann die Zahl der parallel wiedergegebenen Bildinhalte theoretisch sehr groß sein. Das Verfahren wird mittlerweile von der Firma Dolby unter dem Namen Dolby 3D Digital Cinema für die digitale Kinoprojektion eingesetzt⁶².

- KMQ: Dieses Verfahren wurde zu Beginn der 80er Jahre entwickelt und ist benannt nach seinen Entwicklern Dr. Christoph Koschnitzke, Rainer Mehnert und Dr. Peter Quick⁶³. Die beiden Stereohalbbilder werden dabei mit zwei Projektoren übereinander dargestellt; zur Betrachtung wird eine spezielle Prismenbrille benötigt. Die Prismen sind um 180° gegenläufig angeordnet, wodurch der Sehweg der Augen auf das jeweilige Bild umgelenkt wird. Der Betrachter nimmt auf diese Weise beide Stereohalbbilder nebeneinander wahr. Die Methode verursacht keinen Verlust hinsichtlich Helligkeit oder Farben, hat jedoch die Nachteile, dass Stereobilder nur von einem festen Bereich sowie einer bestimmten Entfernung aus betrachtet werden können. Zusätzlich können nur wenige Personen (2-3) gleichzeitig die Wiedergabe verfolgen. Darüber hinaus erzeugt die Prismenbrille eine geringe geometrische Verzerrung im Bild, was die Betrachtungsqualität bei längerem Hinsehen beeinträchtigt⁶⁴.

1.4.2.2 Mit Sehhilfe – aktiv

- Shuttertechnik: Shutterbrillen arbeiten mit zwei LCD-Gläsern (liquid crystal display), deren Lichtdurchlässigkeit gesteuert werden kann. Bei der Wiedergabe werden in schnellem Wechsel linkes und rechtes Stereobild an der gleichen Position dargestellt. Je nach momentaner Ansicht werden dann synchronisiert entweder das linke oder das rechte Brillenglas abgedunkelt⁶⁵. So sieht ein Auge immer nur die entsprechende Ansicht. Um eine unbeeinträchtigte Wahrnehmung beim Betrachter zu gewährleisten, muss die abwechselnde Bilddarstellung mit einer hohen Frequenz erfolgen. Der Mindestwert hierfür liegt bei einer Bildwiederholrate von 120 Hz, somit ist eine Frequenz von 60 Hz pro Ansicht, bzw. Auge gewährleistet⁶⁶. Bei niedrigeren Raten scheint das Bild zu flimmern, was die Betrachtungsqualität erheblich mindert. Vorteile bei dieser Methode sind der Erhalt von Farben und Helligkeit sowie die Tatsache, dass eine herkömmliche Leinwand verwendet werden kann und viele Personen gleichzeitig

61 Vgl. INFITEC GmbH (Hg.) 2010, S. 3.

62 Vgl. Dolby Laboratories Inc. (Hg.).

63 Vgl. Röder 2007, S. 24.

64 Vgl. Röder 2007, S. 25.

65 Vgl. Mahler 2005, S. 281.

66 Vgl. Röder 2007, S. 27.

das Präsentierte betrachten können. Nachteil sind der Helligkeitsverlust und technische Umsetzung der Brillensteuerung, was dazu führt, dass die Brillen vergleichsweise teuer sind.

1.4.2.3 Ohne Sehhilfe

- Parallax-Barriere: Hier wird ein LCD- oder TFT-Display (thin film transistor) durch entsprechenden Umbau modifiziert. Es kommt dabei eine Streifenmaske mit einem sehr feinen Raster senkrecht verlaufender Schlitze zum Einsatz. Im Pixelarray des Displays werden linkes und rechtes Stereobild gleichzeitig dargestellt, in Form von nebeneinander angeordneten senkrecht verlaufenden Streifen, wobei sich immer ein Streifen des linken Bildes mit einem Streifen des rechten Bildes abwechselt. Durch Aufbau und Position der Streifenmaske, welche als Parallax-Barriere bezeichnet wird, werden unterschiedliche Ansichten in Richtung der Augen geleitet⁶⁷. Dadurch, dass jedem Auge eine versetzte Perspektive geliefert wird, kann das Gehirn ein dreidimensional wirkendes Gesamtbild erzeugen⁶⁸. Je nach Displaytechnik muss die Streifenmaske vor oder hinter der eigentlichen Pixelebene platziert werden. Diese Methode erlaubt es, mehr als zwei unterschiedliche Ansichten zu implementieren, um den Betrachtungsraum zu vergrößern und somit mehr Zuschauer zu bedienen. Die Darstellung von mehreren Ansichten bedeutet jedoch auch, dass diese zuvor entweder aufgenommen oder – wenn es sich um CGI-Material handelt – gerendert werden müssen, oder aber dass die zusätzlichen Perspektiven durch Interpolation berechnet werden. Dies funktioniert jedoch oft nur bei bestimmten Bildinhalten, zusätzlich ist die Qualität von Bildern, die durch Interpolation entstehen, unbefriedigend im Vergleich zu einer Originalaufnahme. Bei modernen Parallax-Barrieren-Displays besteht die Barriere aus Flüssigkristallen. Somit ist es möglich, die Barrieren-Funktion zu deaktivieren, so dass auch monoskopische 2D-Inhalte dargestellt werden können (switchable barrier)⁶⁹.

67 Vgl. Mahler 2005, S. 284.

68 Vgl. Lipton 1982, S. 70.

69 Vgl. Schmidt 2009, S. 500.

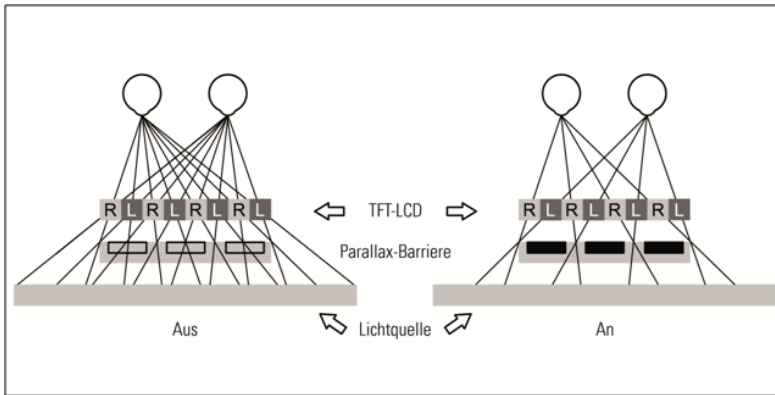


Abbildung 7: Parallax-Barriere

- **Linsenraster:** Grundlage für diese Wiedergabemethode ist ein LCD-Display. Auch hier erfolgt die Darstellung der beiden Stereobilder ähnlich wie bei der Parallax-Barriere-Technik über dünne senkrechte Streifen, die abwechselnd zur linken und rechten Ansicht gehören. Zwischen den Bildstreifen und dem Betrachter befinden sich sehr kleine optische halbzyklindrische Linsen oder Prismen, sog. *Lentikulare*, welche beispielsweise aus Plastik bestehen und in Form einer speziellen Folie auf das Display aufgebracht sind⁷⁰. Ein *Lentikular* überdeckt dabei je einen Pixelstreifen der linken und rechten Ansicht. Durch die Brechungseigenschaften der *Lentikulare* und den Abstand zwischen den Augen des Betrachters sehen je beide Augen das Bild aus einem unterschiedlichen Winkel, wodurch im Gehirn das räumliche Gesamtbild entsteht⁷¹. Diese Methode funktioniert auch für mehr als zwei Ansichten, allerdings nimmt die Bildqualität dabei ab, da sich mehr Ansichten auf die gleiche, unveränderbare Pixelanzahl der Displayfläche verteilen.

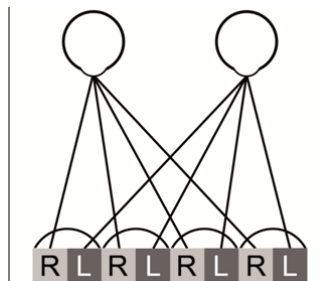


Abbildung 8: Lentikular-Display

⁷⁰ Vgl. Mahler 2005, S. 283.

⁷¹ Vgl. Lipton 1982, S. 71

1.4.2.4 Sonderform Head-Mounted Display

Die Head-Mounted Display-Technik (kurz: HMD) wird hauptsächlich im Zusammenhang mit Virtual-Reality-Umgebungen eingesetzt. Es handelt sich dabei um ein visuelles Ausgabegerät, welches auf dem Kopf getragen wird und Inhalte entweder auf einem oder zwei augennahen Displays wiedergibt oder direkt auf die Netzhäute projiziert (virtual retinal display, kurz: VRD). Durch die unmittelbare Nähe zu den Augen decken diese Bilder im Extremfall das komplette Sichtfeld ab. HMDs mit zwei Displays können auch stereoskopische Bildpaare wiedergeben⁷². HMDs bieten eine Alternative zu konventionellen Bildschirmen, können jedoch logischerweise immer nur von einer Person benutzt werden. Die Geräte werden beispielsweise bei Trainingssimulationen im militärischen und medizinischen Bereich eingesetzt, sowie bei 3D-Computerspielen oder in Forschungseinrichtungen⁷³.

1.5 Aktuelle Situation und Einsatzbereiche

Die Einsatzbereiche von stereoskopischen Verfahren sind ebenso zahlreich wie unterschiedlich. In der Medizin wird die Technik u.a. zur Planung und Durchführung von chirurgischen Eingriffen genutzt⁷⁴, im Zusammenhang mit biochemischen Anwendungsgebieten können Molekülstrukturen dreidimensional dargestellt werden. Die High Resolution Stereo Camera (kurz: HRSC) des Deutschen Zentrums für Luft- und Raumfahrttechnik (kurz: DLR) wird eingesetzt, um Landschaften dreidimensional zu erfassen – unter anderem kam dieses System bei der Marsexpedition der European Space Agency (kurz: ESA) Ende 2003 zum Einsatz⁷⁵. Ein anderer Bereich, in denen stereoskopische Verfahren zum Einsatz kommen, ist die industrielle Bildverarbeitung, wo Stereosysteme beispielsweise zur Qualitätskontrolle von technischen Bauteilen⁷⁶ oder zur Navigation autonomer Fahrzeuge oder Roboter⁷⁷ genutzt werden. Auch im militärischen Bereich wird die Stereoskopie genutzt, neben dem bereits erwähnten Beispiel realistischer Trainingssimulationen sind hier Verfahren zu nennen, die dazu dienen, die Geländebeschaffenheit von Landschaften und Einsatzzonen zu analysieren⁷⁸. Renommierete Forschungseinrichtungen wie das Fraunhofer-Institut für Graphische Datenverarbeitung (kurz: IGD) in Darmstadt bzw. das Fraunhofer Heinrich-Hertz-Institut (kurz: HHI) in Berlin beschäftigen sich mit verschiedenen Projekten zur Thematik 3D-Stereo⁷⁹. Weiterhin laufen derzeit Analysen bezüglich Standards zur Übertragung und Wiedergabe von 3D-Stereoformaten auf Fernsehgeräten in Haushalten. Hierzu hat die Society of Motion Picture and Television Engineers (kurz:

72 Vgl. Schmidt 2009, S. 500.

73 Vgl. Maier 2009, S. 13.

74 Vgl. Trideltity Display Solutions GmbH (Hg.) [1].

75 Vgl. Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR) (Hg.).

76 Vgl. Wöhler 2009, S. 243ff.

77 Vgl. Vincze et al. 2008.

78 Vgl. Trideltity Display Solutions GmbH (Hg.) [2].

79 Vgl. 3sat.online 2009.

SMPTE) 2009 einen entsprechenden Bericht veröffentlicht⁸⁰. Von allen Branchen erfährt im Moment möglicherweise die Unterhaltungsindustrie den stärksten Impuls durch die Stereoskopie. 2009 wurde das internationale Filmfestival in Cannes erstmals mit einem Animationsfilm in 3D-Stereo eröffnet (Oben)⁸¹, James Camerons 3D-Film Avatar ist mittlerweile die kommerziell erfolgreichste Produktion der bisherigen Filmgeschichte⁸² und 3D-Stereo war einer der Themenschwerpunkte der 15. Conference on Animation, Effects, Games and Interactive Media (kurz: FMX) 2010 in Stuttgart⁸³. Auch Hersteller für Computer- und Konsolenspiele widmen sich verstärkt der 3D-Technik. Die Firma BlitzGames Studios aus Leamington Spa in Großbritannien hat beispielsweise während des 3D Entertainment Summits in Los Angeles im Dezember 2008 speziell ausgestattete Spielekonsolen vorgestellt, welche die Wiedergabe von Spielen in 3D-Stereo ermöglichen⁸⁴. Insgesamt darf man gespannt sein, in welche Richtung und in welchen Bereichen sich die Stereoskopie (weiter-)entwickeln wird. Bezogen auf die Filmindustrie glauben Skeptiker nicht an einen Durchbruch der 3D-Technik und verweisen auf die kurzen Booms in der Vergangenheit, in der sich das Verfahren bereits mehrmals nicht am Markt behaupten konnte. Auch wird sich zeigen müssen, ob das Kinopublikum weiterhin mehr Geld für eine 3D-Vorstellung ausgeben wird als für eine klassische monoskopische 2D-Vorstellung. Befürworter argumentieren, dass die Stereoskopie der natürlichen Wahrnehmung des Menschen entspricht und dadurch eine stärkere Immersion ermöglicht wird. Zusätzlich wird oft die fortschreitende Digitalisierung der Filmproduktionskette angeführt, mit der sich viele der früheren Probleme lösen, bzw. abschwächen lassen. Einen wichtigen Beitrag zur breiten Akzeptanz von 3D-Stereo dürfte vermutlich die Entwicklung im Bereich des 3D-fähigen Fernsehens liefern. Jeffrey Katzenberg, Chief Executive Officer von Dreamworks Animation SKG, vertritt die Meinung, dass Filme in absehbarer Zukunft nur noch ausschließlich in 3D-Stereo produziert und konsumiert werden, ähnlich wie der Farbfilm den Schwarz-Weiß-Film ersetzt hat⁸⁵.

80 Vgl. SMPTE (Hg.) 2009.

81 Vgl. Festival de Cannes (Hg.) 2009.

82 Vgl. WELT ONLINE 2010.

83 Vgl. FMX 2010.

84 Vgl. Blitz Games Studios Ltd. (Hg.) 2008.

85 Vgl. Maier Dez. 2008, S. 37.

Literaturverzeichnis

Monografien

- Böhrringer, Joachim; Bühler, Peter; Schlaich, Patrick (2008): Kompendium der Mediengestaltung. [Produktion und Technik] für Digital- und Printmedien. 4., vollst. überarb. u. erw. Aufl. Berlin: Springer (Kompendium der Mediengestaltung Digital und Print).
- Goldstein, Eugen Bruce (2008): Wahrnehmungspsychologie. Der Grundkurs. 7. Aufl. Berlin, Heidelberg: Spektrum Akad. Verl.
- Lipton, Lenny (1982): Foundations of the stereoscopic cinema. A study in depth. New York: Van Nostrand Reinhold.
- Mahler, Gerhard (2005): Die Grundlagen der Fernsehtechnik. Systemtheorie und Technik der Bildübertragung. Berlin, Heidelberg: Springer.
- Ohm, Jens-Rainer; Lüke, Hans Dieter (2005): Signalübertragung. Grundlagen der digitalen und analogen Nachrichtenübertragungssysteme. 9., bearb. Aufl. Berlin, Heidelberg: Springer. (Springer-Lehrbuch).
- Röder, Oliver (2007): Grundlagen der Stereoskopie. Analyse der Aufnahme und Projektion von 3-D Bildern. Saarbrücken: VDM Verl. Müller.
- Schmidt, Ulrich (2009): Professionelle Videotechnik. Grundlagen, Filmtechnik, Fernsehtechnik, Geräte- und Studiotechnik in SD, HD, DI, 3D. 5., aktualisierte und erweiterte Auflage. Berlin, Heidelberg: Springer
- Schreer, Oliver (2005): Stereoanalyse und Bildsynthese. Berlin, Heidelberg: Springer.
- Schröder, Gottfried; Treiber, Hanskarl (2007): Technische Optik. Grundlagen und Anwendungen. 10., erw. Aufl. Würzburg: Vogel (Vogel-Fachbuch).
- Wöhler, Christian (2009): 3D Computer Vision. Efficient Methods and Applications. Berlin: Springer (X).
- Zone, Ray (2005): 3-D filmmakers. Conversations with creators of stereoscopic motion pictures. Lanham Md.: Scarecrow Press (Filmmakers series, no. 119).
- Fachartikel
- Pulfrich, Carl (1922): Die Stereoskopie im Dienste der isochromen und heterochromen Photometrie. In: Die Naturwissenschaften, Jg. 10, H. 25, S. 553–564.
- Vincze, M. et al. (2008): Roboternavigation in Büros und Wohnungen. In: Elektrotechnik und Informationstechnik, Vol. 125, No. 1-2 (2008), S. 25–32.
- Hochschulschriften
- Hottong, Nikolaus (Hg.) (2009): Stereoskope HD-Produktion. Grundlagen Konzepte Testergebnisse. Furtwangen: Fakultät Digitale Medien Hochschule Furtwangen (Schriftenreihe der Fakultät Digitale Medien an der Hochschule Furtwangen).

Internetquellen

- 3sat.online (2009): Die dritte Dimension. 3D-Fernsehen für alle? Online verfü-

- bar unter <http://www.3sat.de/mediathek/frameless.php?url=/hitec/136684/index.html>, zuletzt geprüft am 25.05.2010.
- Blitz Games Studios Ltd. (Hg.) (2008): Blitz Games Studios Debut True 3D Games in World First. Online verfügbar unter http://www.blitzgamesstudios.com/news/Blitz_Games_Studios_Debut_True_3D_Games_in_Word_First/62/, zuletzt geprüft am 25.05.2010.
 - Daro, John (2009): How an industry leader redefines itself for the future. Fotokem: 3D DI. In: Creative COW Magazine, S. 22–25, Online verfügbar unter: <http://magazine.creativecow.net/article/fotokem-3d-di>.
 - Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR) (Hg.): HRSC - die hochauflösende Stereokamera. Online verfügbar unter http://www.dlr.de/mars/desktopdefault.aspx/tabid-290/944_read-1409/, zuletzt geprüft am 25.05.2010.
 - Dolby Laboratories Inc. (Hg.): Dolby 3D Digital Cinema. Online verfügbar unter <http://www.dolby.com/professional/solutions/cinema/3d-digital-cinema.html>, zuletzt geprüft am 25.05.2010.
 - Festival de Cannes (Hg.) (2009): Up. Online verfügbar unter <http://www.festival-cannes.fr/en/archives/ficheFilm/id/10885727/year/2009.html>, zuletzt geprüft am 25.05.2010.
 - FMX (2010): 3D Stereo. Online verfügbar unter <http://www.fmx.de/3d-stereo.E.2221.html>, zuletzt geprüft am 25.05.2010.
 - IRIDAS Dual Stream Products. Herausgegeben von IRIDIAS. Online verfügbar unter <http://www.iridas.com/stereo/>, zuletzt geprüft am 25.05.2010.
 - NVIDIA Corporation (Hg.): 3D Vision. Online verfügbar unter http://www.nvidia.de/object/GeForce_3D_Vision_Main_de.html, zuletzt geprüft am 25.05.2010.
 - Quantel (2008): Stereoscopic 3D. add a new dimension to your business. Herausgegeben von Quantel. Online verfügbar unter <http://www.quantel.com>, zuletzt geprüft am 25.05.2010
 - Seymour, Mike (2008): Art of Digital 3D Stereoscopic Film. Online verfügbar unter <http://www.fxguide.com/article471.html>.
 - Seymour, Mike (2010): The Tech Behind the Tools of Avatar Part 1: Ocula. Online verfügbar unter <http://www.fxguide.com/modules.php?name=News&file=article&sid=594>.
 - Tridelity Display Solutions GmbH (Hg.) [1]: Medizin. Online verfügbar unter <http://www.tridelity.de/Medizin.3dmedicine.0.html>, zuletzt geprüft am 25.05.2010.
 - Tridelity Display Solutions GmbH (Hg.) [2]: Militär. Online verfügbar unter <http://www.tridelity.de/Militaer.military.0.html>, zuletzt geprüft am 25.05.2010.
 - WELTONLINE (2010): „Avatar“ ist der erfolgreichste Film aller Zeiten. Axel Springer AG. Online verfügbar unter <http://www.welt.de/kultur/article5975833/Avatar-ist-der-erfolgreichste-Film-aller-Zeiten.html>, zuletzt geprüft am 25.05.2010.

Sonstige Veröffentlichungen

- Assimilate (2009): SCRATCH and Stereoscopic 3D. Herausgegeben von Assimilate Inc.
- Maier, Florian (Juli/Aug. 2008): Stereoskopische Filmproduktion. Teil 1: 3D-Grundlagen. In: Professional Production, Jg. 22, H. 220, S. 14–18.
- Maier, Florian (Nov. 2008): 3D-Aufnahme-Technologien. In: Professional Production, Jg. 22, H. 23, S. 26–29.
- The Foundry (Hg.) (2010): Product Guide 2010.
- SMPTE (Hg.): Society of Motion Picture and Television Engineerings Inc. (2009): Report of SMPTE Task Force on 3D to the Home. Society of Motion Picture and Television Engineerings Inc. (SMPTE).

Teil 2

Konzeption und Aufbau eines einfachen praxisorientierten S3D-Kamerasystems für Forschungs- und Lehrzwecke

Dirk Hensel, Paul Walter

Abbildungsverzeichnis

Abbildung 1 – Keystone-Effekt	60
Abbildung 2 - HD-RH1, HD-RH1 & CCU, zwei Stingray 125C	67
Abbildung 3 – Objektivsteuerungselemente Preston Cinema Systems HU3, Arri Lens Control, Arri WRS	72
Abbildung 4 – Genlockgeneratoren LynxTechnik Yellowbrick und AJA GEN10 ..	73
Abbildung 5 – Stereorigs Swissrig und RED	74
Abbildung 6 – Feldrekorder Codex Digital Portable, DVC Megacine, Convergent Design FlashXDR	82
Abbildung 7 - Kameraobjektiv mit einfachen Zahnrädern	87
Abbildung 8 - Stereorig	88
Abbildung 9 - Schärfchart (li. Stingray, re. HD-RH1)	93
Abbildung 10 - Helligkeitschart (li. Stingray, re. HD-RH1)	93
Abbildung 11 – Waveform Helligkeit (li. Stingray, re. HD-RH1)	94
Abbildung 12 - Farbchart (li. Stingray, re. HD-RH1)	94
Abbildung 13 - Waveform Farbe (li. Stingray, re. HD-RH1)	95
Abbildung 14- Waveform YCbCr (li. Stingray, re. HD-RH1)	95
Abbildung 15 - Waveform RGB (li. Stingray, re. HD-RH1)	96
Abbildung 16 - Verktorskop (li. Stingray, re. HD-RH1)	96
Abbildung 17 - Testbild (li. Stingray, re. HD-RH1)	100
Abbildung 18 - Lowlight (li. Stingray, re. HD-RH1)	100
Abbildung 19 - Stingray: Außen, Pan, Tilt	101
Abbildung 20 - HD-RH1: Außen, Pan, Tilt	101
Abbildung 21: Kamera AVT GE 1910C	105
Abbildung 22: Objektiv	106
Abbildung 23: Objektiv Schemata	106
Abbildung 23: Aufnahmerechner	107

1 Einleitung

Die Fragestellung für unsere Entwicklungsaufgabe lautete folgendermaßen: Mit welchen auf dem Markt vorhandenen oder selbst zu erstellenden Komponenten kann ein handhabbares, hochauflösendes, digitales Stereovideosystem mit vollem Bedienungs-umfang an der Hochschule Furtwangen realisiert werden?

Das Hauptaugenmerk lag dabei darauf, die praktikabelste Lösung eines stereoskopischen High-Definition-Systems für Lehr- und Forschungszwecke an der Fakultät Digitale Medien zu finden. Bisher wurde schon mit diversen Lösungen experimentiert. Doch wurde hier meist das vorhandene Equipment verwendet, welches nicht optimal für Stereoskopie ausgelegt ist. In diesem Zuge sollen neue Kameras beschafft werden, welche den Anforderungen der Stereoskopie genügen, sich sinnvoll in die Struktur der Hochschule einfügen und gleichermaßen zukunftsorientiert sind. Um dies zu erreichen wird eine umfangreiche Recherche auf dem Gebiet der Kameras durchgeführt und es werden Tests mit verschiedenen Kameras vorgenommen. Diese liefern Vergleichsdaten der verschiedenen Kameraarten und tragen direkt zur Entwicklung eines Kamerakonzeptes bei. Parallel wurde eine neue Montagelösung der Kameras entwickelt, um diese handhabbar einzusetzen. Die technischen und gestalterischen Möglichkeiten des Films werden dadurch erweitert.

Die Verbindung aller Komponenten zu unterschiedlichen Beispielsystemen stellte den Abschluss der Entwicklungsaufgabe dar. Im Folgenden wird nochmals die Vorgehensweise vorgestellt.

Zunächst einmal war es notwendig das Thema in seine einzelnen Teilbereiche aufzuspalten.

Stereoproduktion

Die bisherigen Arbeiten konnten schon einige Erkenntnisse und praktische Erfahrungen zum Thema Stereoproduktion liefern. Auch wurde erkannt, wo Probleme auftreten und das vorhandene Equipment an seine Grenzen gelangt. Aufbauend auf diesen Grundlagen muss nun festgestellt werden, welche Ansätze weiter verfolgt werden

sollen und welche Eigenschaften verbessert werden müssen. Gestalterische und technische Faktoren haben hier einen großen Einfluss auf die Konstruktion des Systems. Auch die Aufzeichnungsmethode, die Speicherung und die verschiedenen Wiedergabeformate sind für die Umsetzung ein wichtiger Anhaltspunkt.

Kamera

Auf dem Markt werden Kameras der verschiedensten Variationen angeboten. Die Gehäuse- und Sensorgröße, das Aufzeichnungsformat und die Stromversorgung sind nur einige wenige Faktoren die bei einem Stereosystem beachtet werden müssen. Je nach Anspruch kann ein individuelles Modell gewählt werden. Die bisher an der Hochschule verfügbaren Geräte konnten die Anforderungen an einen Aufbau für räumliche Videos noch nicht vollständig erfüllen, was starke Einschränkungen für die Produktion bedeutete. Es muss recherchiert werden welche Kameras die geeignetste Lösung für den Betrieb an der Hochschule darstellen. Durch Betriebstests kann dies am geeignetsten festgestellt werden.

Stereorig

Die Montage zweier Kameras spielt eine weitere entscheidende Rolle für die Stereoskopie. Es gibt unterschiedliche Arten von Rigs, welche die Kameras in Position bringen. Durch das Rig kann die stereoskope Aufnahme in ihren wesentlichen Parametern beeinflusst werden. Dies hängt wiederum davon ab, wie genau die Kameras ausgerichtet sind und welche Einstellungen vorgenommen werden können. Desweiteren muss geklärt werden, welche Bauform für die Hochschulausbildung am sinnvollsten ist und wie ein solcher Aufbau realisiert werden kann. Eine Recherche zu den am Markt vorhandenen Konstruktionen soll dazu Anregungen bringen.

Objektivsteuerung

Die synchrone Bedienung von Zoom, Blende und Schärfenregelung beider Kameraobjektive stellt eine besondere Herausforderung dar. Um während der Aufnahme einen der Parameter zu verstellen, müssen sich beide Objektive genau gleich verhalten. Der räumliche Eindruck wird zerstört, wenn sich die Schärfe asynchron verhält, oder der Bildausschnitt sich Rechts und Links unterscheiden würde. Die manuelle Steuerung stellt die gefühlvollste und kostengünstigste Lösung dar und soll deshalb bevorzugt werden. Da eine feste Verbindung der Objektive aufgrund der variablen Stereobasis nicht zum gewünschten Ergebnis führen würde, muss eine sinnvolle, präzise Steuerung konzipiert werden.

Objektiv

Da die meisten in Frage kommenden Kameras standardmäßig ohne Objektiv geliefert werden, muss auch auf diesen Punkt weiter eingegangen werden. Die Frage der Optik ist für ein Stereosystem noch bedeutender als für den herkömmlichen Film. Dies hängt unter anderem mit der Fertigungsqualität, dem Brennweitenbereich und der Auflösungsqualität zusammen. Erörtert werden muss, welche konkreten Objektive

für eine Verwendung an der Hochschule geeignet wären. Dabei sollen alle wichtigen technischen und gestalterischen Einflussfaktoren berücksichtigt werden.

2 Technische und gestalterische Parameter

2.1 Stereobasis

Die Stereobasis beschreibt den horizontalen Abstand beider Objektivmittelpunkte voneinander. Sie kann als Übersetzung des Augenabstandes auf eine technische Ebene gesehen werden. Außerdem zählt sie zu den wichtigsten Parametern in einem Stereosystem. Der angestrebte Idealabstand paralleler Kameras beträgt etwa 6,5cm, was dem durchschnittlichen, menschlichen Augenabstand entspricht. Für diesen Wert, ist die Abbildung der räumlichen Szene für den Betrachter am natürlichsten, weil sie genau der Anatomie nachempfunden ist. Wird die Stereobasis verändert, hat das zur Folge, dass andere Kameraeinstellungen mit ausreichender Tiefenwirkung möglich werden, aber auch mit verschiedenen Eigenheiten gerechnet werden muss¹. Beispielsweise werden Makroaufnahmen mit einer kleineren Stereobasis vorgenommen, um die entstehende Parallaxe nicht zu groß werden zu lassen, da die Objekte sehr nah an der Linse positioniert sind. Panoramaaufnahmen können hingegen mit einem größeren Abstand vorgenommen werden, um die Tiefenwirkung von weit entfernten Szenen aufrecht zu erhalten. Bei dieser Manipulation der Stereobasis können jedoch Effekte wie Mikropsie oder Gigantismus auftreten². Die Bilder erscheinen dann zwar räumlich, können jedoch je nach Intensität unnatürlich groß bzw. klein wirken.

Nicht nur die horizontale, sondern auch die vertikale Ausrichtung der Kameras, die Verschiebung nach vorn oder Hinten und der Konvergenzwinkel sind bei der Montage zu beachten. Es ist von großer Bedeutung, dass die Kameras auf der exakt selben Höhe und Entfernung zur Szene stehen. Andernfalls könnten die stereoskopischen Halbbilder nicht zu einem korrekten dreidimensionalen Bild zusammengefügt werden. Auch der Montagewinkel der Kameras ist von großer Bedeutung um ein angenehmes Raumbild zu erzeugen. Die Objektive dürfen von einander weg zeigen, da in diesem Fall Divergenzen entstehen, welche für den Menschen unnatürlich sind³.

2.2 Brennweite

Der menschliche Sehapparat hat eine feste Brennweite von etwa 25mm.⁴ Beim Film wird die Brennweite aus gestalterischen und technischen Gründen oft verändert. Bei der Verwendung in der Stereoskopie kann dieselbe Vorgehensweise beibehalten werden. Allerdings ergeben sich einige Unterschiede zum herkömmlichen Film, welche beachtet werden müssen.

Am besten funktioniert die räumliche Wirkung bei kurzen Entfernungen und kurzen Brennweiten. Lange Brennweiten neigen dazu nicht nur den aufgenommenen Raum

1 Vgl. Oliver Röder, Röder 2007, S. 12

2 Vgl. Ray Zone 2005, S. 26,64,69

3 Vgl. Oliver Röder, Röder 2007, S. 21

4 Vgl. Lipton 1982, S. 62

sondern auch das Volumen von Objekten und Personen zu stauchen. Der aus langen Brennweiten resultierende, sogenannte Cardboarding-Effekt hat zur Folge, dass alle Objekte flach wirken und durch diesen Eindruck ein unnatürliches Bild entsteht. Der Grenzwert für das räumliche Sehen kann zwar mit diversen Formeln berechnet werden, trifft aber nicht für alle Menschen zu und ist deshalb nicht zuverlässig.⁵

Die Normalbrennweite bietet dem Betrachter die natürlichste dreidimensionale Darstellung ohne Verzerrung. Sie hängt von der Größe des Kamerasensors ab. In der Fotografie, die in diesem Fall identisch zum Film ist, wird die Normalbrennweite als die dem Film entsprechende Diagonale bezeichnet. Im Fall der Digitalkameras entspricht dies der Diagonalen des Sensors.⁶ Dabei ist zu beachten, dass die Zollangaben nur als Referenzierung auf die lichtempfindliche Fläche ehemaliger Aufnahmeröhren zu betrachten sind. Die exakten Werte lauten für 2/3“-Sensoren daher 16.93316mm für 1/2“-Sensoren 12,7mm und für 1/3“-Sensoren 8.46658mm. Allgemeine Richtwerte liegen für einen 2/3“ Sensor bei 14-20mm, für 1/2“ bei 10-14mm und für 1/3“ bei 7-10mm. Allerdings hängt die Normalbrennweite auch stark vom persönlichen Empfinden und Stil ab.⁷ Zusätzlich verringert sich mit zunehmender Brennweite und Abstand auch die Tiefenschärfe, was unter bestimmten Umständen nicht erwünscht ist. Speziell bei der Stereoskopie muss zusätzlich beachtet werden, dass je höher die Brennweite ist, auch die Synchronisation der Objektive genauer arbeiten muss.

Kurze Brennweiten können dagegen eine Tonnenverzerrung des Bildes zur Folge haben. Auch kann es sein, dass die Parallaxe beider Stereohalbbilder zu groß wird. Letzteres kann durch verkleinern der Stereobasis korrigiert werden, allerdings stellt z.B. im Side by Side-Betrieb die Größe der Kameragehäuse und Objektive eine physikalische Untergrenze für die Stereobasis dar.

2.3 Scheinfensterweite

Bei der stereoskopischen Filmproduktion stellt die Scheinfensterweite eine wichtige Variable dar. Sie legt die räumliche Ebene fest, und somit auch welche Abstände die Akteure und Objekte in der Szene einnehmen dürfen, um eine ideale räumliche Wirkung zu erzielen (Komfortzone, Tiefenbereichsweite/TBW). Umgekehrt kann auf diese Weise ermittelt werden, wo die Kameras zur Szene positioniert werden müssen. Die SFW wird zunächst ermittelt durch Multiplikation von Stereobasis in Zentimetern mit Brennweite in Millimetern. Hierbei wird deutlich wie die verschiedenen Parameter aufeinander Einfluss haben. Es ist wichtig zu beachten, dass die Parallaxe zwischen den Stereohalbbildern nicht zu groß wird. (Siehe Kapitel 2.4.4)

2.4 Kadrierung

Die Wahl des Bildausschnittes beziehungsweise der Einstellungsgröße hat ebenfalls

5 Vgl. Lipton 1982, S. 109–110

6 Vgl. PENTAX Europe GmbH

7 Vgl. Movie-College.de 28.07.2008

einen Einfluss auf die Parameter Brennweite und Stereobasis. Eine grobe Faustregel ist die 1/30 Regel. Diese besagt, dass der Unterschied von Nahpunkt und Fernpunkt, die sogenannte Deviation, nicht größer als 1/30 der Bildbreite betragen darf, da andernfalls der horizontale Versatz der Stereohalbbilder zu groß wird. Für eine Stereobasis von 65mm multipliziert mit einer Brennweite von 35mm würde das einen Mindestabstand der Kamera vom Nahpunkt von 2,27m bedeuten.⁸ Soll dieser Abstand verringert werden, kann die Brennweite oder die Stereobasis angepasst werden. Vor allem bei nahen Einstellungen, wie Großaufnahme und Details, kann die Stereobasis ein Problem darstellen und somit die gewünschte Einstellungsgröße nicht realisiert werden. Kann die Stereobasis in solch einer Situation nicht verkleinert werden, ist der Abstand von linkem und rechtem Stereohalbbild zu groß und kann aus diesem Grund später schlecht zu einem korrekten Raumbild zusammengefügt werden. Verringert man jedoch den Abstand muss wiederum das zuvor erwähnte Gigantismus-Phänomen beachtet werden. Bei Supertotalen oder Totalen hingegen geht die Räumlichkeit des Bildes verloren, da die Abstände zu den Objekten sehr groß sind. Für den Menschen entsteht ab etwa 50 Metern Entfernung von Betrachter zu Objekt kein räumlicher Eindruck mehr.⁹ Der angenommene Abstand zu den entferntesten Objekten im Spielfilm, ausgenommen Landschaftspanoramen, bei welchen der Abstand auch mehrere Kilometer betragen kann, liegt bei etwa 100 bis 200 Metern. Bei diesem Extrem kann die Stereobasis wiederum zum Ausgleich verwendet werden. Man vergrößert sie um einen dreidimensionalen Eindruck beizubehalten, sollte jedoch das erwähnte Risiko der Mikropsie in Betracht ziehen.

2.5 Tiefenschärfe

Je nach gestalterischer Entscheidung kann die Ausprägung der Tiefenschärfe einen entscheidenden Einfluss auf die Produktion haben. Es gibt zwei verschiedene Herangehensweisen an die Erstellung eines stereoskopischen Films.¹⁰ Einerseits kann die Tiefenschärfe so groß wie möglich gehalten werden, um dem Publikum selbst die Entscheidung zu überlassen, welchen Bereich des Bildes es betrachten möchte. In diesem Fall sollte eine parallele Kameraausrichtung gewählt werden. Durch die Einstellung der Blende, die möglichst geschlossen sein sollte, um eine hohe Tiefenschärfe zu erreichen, ist auch mehr Licht notwendig. Dies kann an einigen Drehorten oder zu bestimmten Tageszeiten problematisch werden. Zumindest bedeutet es aber einen höheren Aufwand für die Beleuchtung der Szene.

Andererseits kann die Tiefenschärfe wie beim zweidimensionalen Film gering gehalten werden, um die Aufmerksamkeit des Betrachters zu lenken. Hierzu kann eine konvergente Kamerastellung verwendet werden, die direkt auf das darzustellende Objekt fokussiert. Dabei können im Hintergrund hohe Disparitäten auftreten, welche den

⁸ Vgl. Oliver Röder, Röder 2007, S. 20

⁹ Vgl. Oliver Röder, Röder 2007, S. 9

¹⁰ Vgl. Eric Schumacher 2008, S. 56

räumlichen Eindruck zerstören können. Somit ist eine hohe Tiefenschärfe bei dieser Variante eher hinderlich.

2.6 Parallel oder Konvergent

Das räumliche Sehen beim Menschen entsteht, wie schon zuvor erkannt, durch die Betrachtung einer Szene von zwei horizontal verschobenen Standpunkten, den Augen. Durch das Verarbeiten der Informationen entsteht ein dreidimensionales Bild. Die Wahrnehmung der Distanz, betrachteter Gegenstände, wird dabei durch viele Faktoren beeinflusst, wie in der Einleitung dieses Arbeitspapiers schon erörtert wurde. Ein Faktor ist die Akkommodation. Hierbei handelt es sich um das Verhalten der Zirkularmuskeln der Augen, um Objekte in bestimmter Entfernung zu fokussieren und die dadurch vom Gehirn interpretierte Distanz von Objekten. Ein weiterer Faktor ist die Konvergenz. Diese beschreibt die Bewegung der Augen zu Nase hin. Dies ist zur Betrachtung naher Gegenstände ebenfalls notwendig und trägt zur Einschätzung der Distanz bei.¹¹

Im Fall einer konvergenten Kamerastellung wird versucht, das Verhalten der Augen technisch umzusetzen. Für nahe Objekte wird in diesem Fall eine stärker konvergierende Kameraeinstellung gewählt, weiter entfernte Objekte werden mit einer nahezu parallelen Einstellung aufgezeichnet. Der Schnittpunkt der Kameraachsen muss dabei genau auf dem fixierten Objekt liegen.¹² Dadurch ergibt sich für den Film das Problem, dass die konvergierenden Kameras permanent richtig ausgerichtet sein müssen, um ein angenehmes Bilderlebnis zu erreichen. Da sich jedoch die Personen und Kameras im Film bewegen, sollte die Kameraausrichtung kontinuierlich nachjustiert werden. Eine Lösung wäre in diesem Fall, die Winkel mechanisch für jede Szene einzustellen. Eine weitere Möglichkeit, die auch James Cameron verwendet, ist die elektronische Kopplung von Schärfe und Konvergenz. Diese wird dann entweder automatisch oder von einem Focuspuller bedient.

Die Verwendung von Konvergenzen ist unter 3D-Interessierten bisher sehr umstritten. Dies ist vor allem darauf zurückzuführen, dass verschiedene Probleme entstehen können. Die Literatur besagt, dass ein konvergentes System in drei Bereiche eingeteilt werden kann. Einen vorderen, einen mittleren und einen hinteren Bereich. Im vorderen Bereich, also näher zur Kamera, ist es schwierig die Stereohalbbilder zu verarbeiten, da der Konvergenzwinkel zu groß ist. Der mittlere Bereich kann wiederum in zwei Bereiche eingeteilt werden. In ihm schneiden sich die Kameraachsen, somit ergibt sich der Bereich vor und hinter dem Schnittpunkt bzw. des Scheinfensters. Im Bereich vor dem Schnittpunkt werden Bilder in verarbeitbaren Grenzen vor der Leinwand abgebildet, dahinter befindet sich der räumliche Bereich hinter dem Scheinfenster, welcher optimal verarbeitet werden kann. Der letzte Abschnitt beschreibt die

11 Vgl. Eric Schumacher 2008, S. 39–41

12 Vgl. Eric Schumacher 2008, S. 55–56

Zone, in welcher die Divergenz einen Gesamtwinkel von 1° überschreitet und sich somit nicht mehr innerhalb der zulässigen Grenzen befindet.¹³

Durch die Fokussierung auf einen bestimmten Punkt in der Szene entstehen im Vorder- und Hintergrund starke Querdisparitäten. Diese zerstören den räumlichen Eindruck und führen zu Unwohlsein beim Betrachter.¹⁴ Um den störenden Effekt zu vermeiden müssen Systeme mit kleinen vorderen und hinteren Tiefenschärfebereichen verwendet werden, wie sie auch beim herkömmlichen Kinofilm üblich sind. Der unscharfe Vorder- und Hintergrund „vertreibt“ eher das saccadisch über das Bild wandernde Auge (quasi-Maskierung) und bietet dem Betrachter folglich weniger Möglichkeiten der falschen Wahrnehmung.

Ebenfalls negativ bei konvergenten Kameras wirkt sich der Keystone-Effekt an den Bildrändern aus, was übersetzt soviel wie Schrägprojektion bedeutet.

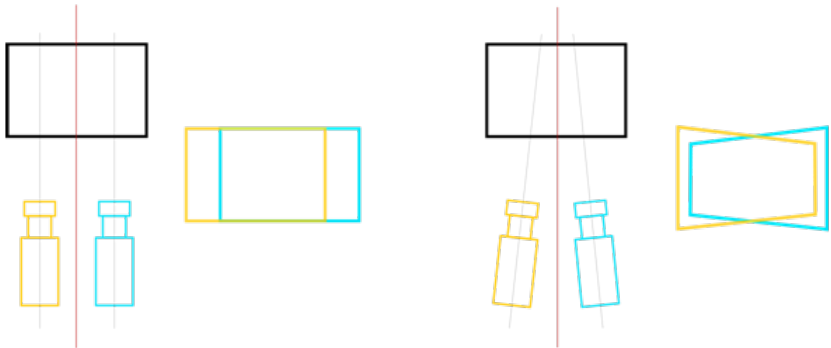


Abbildung 1 – Keystone-Effekt

Dieses Phänomen beschreibt die perspektivische Verzerrung von Flächen bei einer konvergenten Kamerastellung. Werden zwei Kameras in einem bestimmten Winkel zueinander verdreht, ergeben sich unterschiedliche Abstände für die jeweilige Kamera von linkem Bildrand zu Objekt und von rechtem Bildrand zu Objekt. Es entstehen also Verzerrungen der Szene und sowohl horizontale als auch vertikale Parallaxe¹⁵(siehe Abb. 1). Die resultierenden Bilder müssen, bevor sie als Raumbild montiert werden, einen De-Keystoneing unterzogen werden.

¹³ Vgl. Lipton 1982, S. 102

¹⁴ Vgl. Eric Schumacher 2008, S. 55

¹⁵ Vgl. Andrew Woods Feb. 1993

2.7 Kamerafunktionen

Kameras, die in einem stereoskopischen Verbund betrieben werden sollen, müssen einige grundlegende gestalterische Einstellmöglichkeiten besitzen. Zu ihnen gehören vor allem manuelle Bedienung von Blende, Fokus, Brennweite und Verschluss. Diese müssen für jede Kamera identisch nachvollziehbar sein, um korrekte Stereohalbbilder zu erzeugen. Wird ein Wert an einer Kamera verändert, muss auch der Wert der anderen Kamera geändert werden. Besonders während der Aufnahme dürfen sich die Einstellungen nicht unterscheiden. Sollte eine Brennweitenänderung während der Aufnahme vorgenommen werden, muss diese vollständig synchron mit beiden Objektiven vollzogen werden. Um Farbabweichungen innerhalb der Stereohalbbilder zu vermeiden muss auch die Einstellung des Weißabgleich manuell möglich sein. Eine Korrektur in der Postproduktion wäre zwar durchführbar, jedoch vergleichsweise zeit- und arbeitsintensiv. Zuletzt muss noch die Synchronisation der Bilder gewährleistet sein. Zu diesem Zweck können Master-Slave-Schaltungen oder zentrale Genlock-Geräte genutzt werden. Diese sorgen für eine zeitgleiche Aufnahme der Bilder und deren Übertragung. So wird sichergestellt, dass keine zeitliche Verschiebung zwischen den Stereohalbbildern auftritt, was wiederum den räumlichen Eindruck bei der Betrachtung zerstören könnte.

3 Ausgangssituation

3.1 Vorhandenes Equipment

Die bisherigen 3D-Filme und Tests an der Hochschule wurden mit zwei JVC GY-HD251 Kameras auf einer Stereoschiene durchgeführt (ausführlich dargestellt im Arbeitspapier Nr.5). Diese werden mittels HD-SDI Verbindungen an den Aufzeichnungsrechner angeschlossen. Die integrierten Blackmagic Decklink HD-Extreme Video Capture Karten sorgen für eine adäquate Übertragung. Je eine Karte wird einer Kamera zugewiesen, was jeweils einen unkomprimierten Datenstrom mit einer 8bit Quantisierung, 4:2:2 Chromasubsampling und einem Format von 720p50 zulässt. Die von Volker Stephan entwickelte Direct Show-Anwendung schreibt die Informationen dann in einen .AVI Container. Die Streams werden separat für jede Seite auf die Festplatte aufgezeichnet und können mittels einer Ordnerstruktur organisiert werden. Zur Synchronisation beider Kameras dient die Verbindung mittels Koaxialkabel. Dieses übermittelt den Timecode für den Master-Slave-Betrieb. Zusätzlich kann gegebenenfalls eine Regieklappe zum Einsatz kommen. Die Bänder in der Kamera werden parallel zur Direktübertragung genutzt, um eine Sicherungskopie der Aufnahmen zu erhalten. Diese Kopien erreichen - aufgrund der verbauten HDV-Codecs - allerdings nicht dieselbe Qualität wie das vom Rechner aufgezeichnete Signal.

3.2 Vorausgehende Erkenntnisse

Die Einschränkung der Stereobasis in o.g. Anordnung auf minimal 13,5cm, hinderte Eric Schumacher und Anja Frank bei ihren Produktionen Großaufnahmen und Detailaufnahmen im Film zu integrieren. Sie sind allerdings für Filmgestaltung und Bildsprache wichtig. Damit zusammenhängend kann auch die Brennweite der JVC Kameras nur in einem Bereich von 30-50mm genutzt werden. Wobei die Eingrenzung für lange Brennweiten durch das Cardboarding-Phänomen begründet ist. Die Stereobasis und die Scheinfensterweite müssen von Hand abgemessen werden. Eine exakte Einstellung des Kamerawinkels auf der Stereoschiene erfolgt mit verschiedenen Hilfsmitteln. Eine weitere einschlägige Erkenntnis ist, dass der Aufzeichnungsrechner mit seinem Gewicht und seinen Abmessungen, sowie der Stromversorgung die Mobilität sehr einschränkt. Eric Schumacher löste dieses Problem, indem er alle Komponenten auf einen Dolly montierte und so die Bewegungsfreiheit und den Bedienungskomfort erhöhte. Die Synchronisation der Kameras über die Timecode-Schnittstelle funktioniert gut, was eine Weiterführung dieses Konzeptes erstrebenswert macht. Für den Aufnahmeserver ist außerdem aufgefallen, dass die Schreibgeschwindigkeit nur kleine HD-Videoformate verarbeiten kann, eine Verbesserung auf ein 1080p Format also nicht durchgeführt werden könnte. Dies wird auch von den Blackmagic-Karten verhindert, welche maximal ein 1080i Video unterstützen. Der mitgelieferte Geschwindigkeitstest der Video-Karten zeigt, dass vom eingebauten Festplatten-RAID eine Schreibgeschwindigkeit von etwa 200MB/s erreicht wird. Für die Einstellungen der

Kamera muss eine weitere Einschränkung bezüglich des Objektivs gemacht werden. Dieses kann nicht bei laufender Aufnahme bedient werden, da eine Synchronisation mit dem Objektiv der zweiten Kamera nicht gewährleistet ist. So werden Schärfenverlagerungen und Zooms für eine Produktion ausgeschlossen.

3.3 Notwendiges Equipment

Es lässt sich also ableiten, dass mehrere Veränderungen vorgenommen werden muss, um die bisherigen Problempunkte zu verbessern. Um eine Stereobasis von 6,5cm zu erreichen, gibt es nur zwei Möglichkeiten. Es können entweder kleinere Kameras eingesetzt oder eine andere Rig-Variante verwendet werden. Im weiteren Verlauf der Arbeit werden die verschiedenen Aspekte der Rigs und Kameras tiefer behandelt. Um ein voll einsatzfähiges System zu erhalten ist außerdem die Entwicklung einer Objektivsteuerung notwendig, welche die Objektive der Kameras synchron einstellen kann. Dies kann auf elektronischem oder mechanischem Weg geschehen, wobei hier die Größe der Objektive und ihre Steuerung einen entscheidenden Einfluss haben können. Das Problem der Mobilität kann nur durch einen Austausch des Aufzeichnungsrechners gegen ein handlicheres Gerät gelöst werden, welches eine höhere Leistung erzielt. Zudem sollte es so gut wie möglich mit einem Akku betrieben werden können, um sich von einer festen Stromverbindung abzukapseln. Dasselbe gilt auch für den Betrieb aller übrigen verwendeten Geräte. Ein Monitor zur Kontrolle des Bildes und der Kameraeinstellungen sollte ebenfalls in den Aufbau eingeplant werden, da einige der in Frage kommenden Mikrokameras über keine alternative Vorschaumöglichkeit verfügen. Es besteht auch die Möglichkeit, dass die Kameras keinen Timecode abgeben und somit nicht im Master-Slave-Modus geschaltet werden können. Daher muss eine andere Lösung der Synchronisation gefunden werden. Denkbar wäre ein Timecode- oder Genlockgenerator.

4 Technische Voraussetzungen

4.1 Geeignete Kameras

4.1.1 A-Cam DII

Ein hochwertiges Beispiel für eine unabhängige Kamera ist die A-Cam DII der Firma Ikonoskop. Das Gerät hat, wie die üblichen Consumer Kameras, ein eingebautes Aufzeichnungssystem und eine interne Stromversorgung. Da die A-Cam DII in hochauflösenden Formaten und ohne Kompression aufzeichnet, handelt es sich hierbei um eine Flash-Speicherkarte und nicht um eine DVD oder ein Magnetband wie bei vielen Consumer-Geräten. Sie verfügt über einen Sucher, mit welchem das Bild und die Kameraeinstellungen überprüft werden können. Der Aufbau des Geräts ist also mit dem einer Einsteiger-Kamera zu vergleichen, die Qualität soll jedoch professionellen Standards gerecht werden. Es muss außerdem berücksichtigt werden, dass die Objektive auswechselbar und die Einstellungsmöglichkeiten ebenfalls mit einer professionellen Kamera vergleichbar sind. Die Abmessungen befinden sich etwas über den Anforderungen des stereoskopischen Betriebs, sind jedoch noch ausreichend. Die A-Cam DII ist somit eine Ausnahme im Bereich der eigenständigen Kameras.

4.1.2 GY-HM100E

Die JVC GY-HM100E ist ein weiteres Beispiel für eine Stand Alone Kamera. Sie ist derzeit die kleinste Prosumer-Kamera auf dem Markt. Die Einstellungsmöglichkeiten umfassen alle notwendigen Faktoren und auch die Stromversorgung ist integriert. Nachteilig bemerkbar machen sich aber vor allem die Abmessungen, welche einen Stereobetrieb nur mit einem Mirror-Rig zulassen, und die kamerainterne Kompression der Videodaten. Das MPEG2 Long GOP oder Quicktime Format mit einer Datenrate von ca. 35Mb/s lässt die Qualität des Signals deutlich sinken. Ein weiteres Problem stellt die Synchronisation zweier Kameras dar. Mit der GY-HM100E ist, aufgrund mangelnder Anschlüsse, keine Synchronisation durchführbar. Diese Fakten sind exemplarisch für die meisten Prosumer-Kameras, ausgenommen Sony und Canon Modelle. Diese verfügen oft über einen sogenannten LANC-Anschluss. Dieser ermöglicht eine einfache Synchronisation mehrerer Kameras per Funk- oder Kabelsteuerung.

4.1.3 Kameras mit abgesetzten CCU

Diese Kameras zeichnen die Bilder nicht direkt in der Kamera auf sondern reichen sie nur an ein Aufnahmegerät weiter. Zwischen Aufnahme und Kamera muss eine sogenannte Camera Control Unit (CCU) angeschlossen werden. Über sie werden die nötigen Einstellungen der Kamera vorgenommen. In manchen Fällen kann auch das Objektiv einer Kamera mit der CCU gesteuert werden um auf diesem Weg Brennweite, Blende und Schärfe zu beeinflussen. Bei dieser Art von Kamera muss eine externe Stromversorgung gewährleistet sein oder in einigen Fällen eine Akkulösung entwickelt werden. Im Bereich der Aufzeichnungsgeräte gibt es verschiedene Lö-

sungen die sich in ihrer Leistung und Handlichkeit unterscheiden. Verwendet werden Bandgeräte, Festplattenrecorder oder RAID-Verbünde mit bis zu 24 Festplatten, die 4:2:2 oder 4:4:4 Datenströme handhaben können. Durch die Auslagerung der Aufzeichnungstechnik und der Steuerung können hier sehr kleine Kameras produziert werden, die gleichzeitig eine sehr hohe Bildqualität liefern.

4.1.4 HD-RH1

Als Beispiel für eine Kamera mit Kontrolleinheit wäre die HD-RH1 von Iconix zu erwähnen. Sie verfügt über einen ausgesprochen kleinen Kamerakopf, an welchen das Objektiv angebracht wird. Eine Kabelverbindung überträgt die Videodaten permanent an die Kontrolleinheit. An ihr können die verschiedenen Formate verändert, der Weißabgleich und alle anderen relevanten Einstellungen vorgenommen werden. Der Unterschied zu den unabhängigen Kameras besteht in der externen Speicherung und Stromversorgung. Für die HD-RH1 ist eines der in Kapitel 4.1.2 genannten Aufzeichnungskomponenten notwendig. Da die Kamera eindeutig im professionellen Bereich angesiedelt ist, werden auch die entsprechenden unkomprimierten Formate für hochauflösendes Video unterstützt. Dies schafft einerseits großen Spielraum für die Bildqualität, sorgt andererseits aber auch für hohe Datenraten. Dies wiederum setzt ein adäquates Aufnahmegerät voraus.

Auch die Stromversorgung wird serienmäßig über ein Netzteil und nicht durch einen Akku gewährleistet, obgleich ein Betrieb mit Batterien denkbar wäre. Durch die Kontrolleinheit bekäme ein Stereosystem zwei zusätzliche Komponenten, da jede Kamera eine eigenständige CCU benötigt, die zusätzlich untergebracht und getragen werden müssen.

4.1.5 Kameras ohne CCU

Diese Kameras unterscheiden sich von den zuvor genannten Bauformen dadurch, dass sie direkt von einem Computer gesteuert werden und meist auch auf diesen aufzeichnen. Die Einstellungen werden auf einem Monitor oder dem Computerbildschirm angezeigt und kontrolliert. Hier ist, wie bei den CCU gesteuerten Varianten, eine sehr kleine Bauform möglich, was für die Stereoskopie sehr von Vorteil ist. Die Stromversorgung, die zusammen mit der Aufzeichnung für die größte Beschränkung in der Handlichkeit und der Mobilität darstellt, kann hier durch den Computer direkt oder durch eine externe Stromversorgung gewährleistet werden.

4.1.6 Stingray 125c und Modula Junior

Bei der Stingray 125C handelt es sich um eine Industriekamera des Herstellers Allied Vision Technologies. Sie soll als Beispiel für eine computergesteuerte Kamera dienen. Angeschlossen wird das Gerät per Firewire800 und auch über diese Verbindung mit Strom versorgt. Andere, von Kameras in diesem Segment verwendete Schnittstellen, sind Ethernet oder SDI. Alle Kameraeinstellungen werden über die Software der Kamera vorgenommen, was eine zusätzliche Kontrolleinheit überflüssig macht. Im Fall

der Stingray kann der Computerbildschirm zur Bildkontrolle verwendet werden. Je nach Monitor können aber Probleme bei der Farbkontrolle und der Schärfe des Bildes auftreten. Abhängig von der eingebauten Hardware ist auch eine leichte Latenz des Videos über den Rechner zu erwarten. Kameras mit SDI-Verbindung wie die Modula Junior von Easylook System, benötigen in der Regel einen eigenständigen Monitor. Diese mini Broadcast-Kamera benötigt ebenfalls keine CCU, verfügt jedoch über eine sogenannte Breakout-Box, über die das Videosignal weiterübertragen wird. Die Box liefert außerdem die notwendige Betriebsspannung per Akku oder Netzteil. Aufgenommen wird der Videodatenstrom bei der Stingray direkt auf dem Rechner. Bei einer Ethernet Verbindung wäre der Ablauf identisch. Auch hier stellt der SDI-Anschluss die Ausnahme dar. Es können neben einem PC alle Aufzeichnungsgeräte, wie Festplattenrekorder oder Bandgeräte, mit passendem Anschluss verwendet werden. Die Einstellungen für den Betrieb der Modula werden extern vorgenommen. Es werden aus diesem Grund Schnittstellen wie Ethernet oder RS485 bereitgestellt. An diese können professionelle Bedienpulte zur Fernsteuerung oder ein Laptop angeschlossen werden. Die Einstellungen können mit jedem gängigen Webbrowser vorgenommen werden. Für den Hochschulbetrieb bedeutet dies, dass in jedem Fall ein Rechner für diese Art von Kamera verwendet werden muss und der Komponentenaufwand sogar noch höher ist als bei Kameras mit CCU.

4.1.7 Kameravorraussetzungen

Um in einem Stereosystem an der Hochschule zu funktionieren, müssen die Kameras einige Kriterien erfüllen. Drei Kategorien werden unterschieden und bestimmen die Wichtigkeit für die Hochschule. Die erste Kategorie legt fest, welche Anforderungen zwingend von den Kameras erfüllt werden müssen. Die Abmessung der Kamera, das Aufzeichnungsformat und die manuelle Bedienung sind unter Anderem festgelegte Kriterien. Werden diese nicht erfüllt, ist die Kamera nicht für das System einsetzbar. In der zweiten Kategorie werden Eigenschaften aufgeführt, die den Betrieb an der Hochschule verbessern. Diese wären unter anderem Robustheit, schnelle Betriebsbereitschaft oder Batteriebetrieb. In der letzten Kategorie sind zusätzliche Eigenschaften, welche sich zwar verbessernd auswirken, jedoch nicht unbedingt erfüllt werden müssen, festgelegt. Eine vollständige Liste ist im Anhang der Arbeit zu finden.

4.1.8 Versuchsauswahl

Für Messungen und Tests an der Hochschule wurden einige Kameras bestimmt. Diese wurden bewusst so gewählt, dass ein breites Spektrum an Kameravarianten verfügbar ist. So soll herausgefiltert werden, welche Eigenheiten, Vor- und Nachteile die jeweilige Bauform mit sich bringt. So wurden Geräte aus jedem der oben genannten Bereiche ausgewählt, mit Ausnahme der unabhängigen Kameras, da bisher auf diesem Gebiet keine einsetzbare Kamera vorhanden ist. Die Testkameras entsprechen den schon zuvor genannten Beispielkameras Iconix HD-RH1 und AVT Stingray 125C. Für die Kategorie der unabhängigen Kameras war die Ikonoskop a-cam dII vorgesehen. Die Durchführung war aufgrund der immer noch andauernden Entwicklung

des Geräts, nicht möglich. Auch die geplante Kamera Modula Junior von Easylook System konnte, trotz vorheriger Zusage, wegen intensivem Produktionseinsatz von Seiten des Herstellers, nicht getestet werden.



Abbildung 2 - HD-RH1, HD-RH1 & CCU, zwei Stingray 125C

4.2 Objektiv

4.2.1 Anschluss

Um ein Objektiv zu ermitteln, das sich sinnvoll in ein Stereosystem einliedert, müssen sehr viele Faktoren beachtet werden. Ein grundlegendes Merkmal stellt der Kameraanschluss dar. Für Video und Filmkameras gibt es verschiedene Standards, welche die Befestigung des Objektivs beschreiben. Gängige Formate sind PL-, B4-, C- und CS-Mount. PL- und B4-Mounts finden hauptsächlich bei den großen Filmkameras oder dem ENG (Electronic News Gathering) Anwendung. Dagegen werden bei den relevanten Videokameras, die für das System ausgesucht wurden, hauptsächlich C-Mounts verwendet. Dies ist ein spezielles Gewinde im Zollformat, das eine geringe Baugröße ermöglicht. Der CS-Mount ist lediglich eine Weiterentwicklung des C-Mount und verspricht noch kleiner und unkomplizierter zu sein. Die beiden Varianten sind mittels eines Adapterrings untereinander kompatibel.

4.2.2 Sensorgröße

Der Sensor der Kamera kann verschiedene Abmessungen haben. Die am häufigsten vertretenen Sensorgrößen sind 1“, 2/3“, 1/2“ und 1/3“ (die Zollangaben referenzieren auf vergleichbare Durchmesser bei früheren Röhrenkameras). Je nach Abmessung muss das Objektiv folglich größere oder kleinere Linsendurchmesser besitzen, um die Kamera optimal zu bedienen. Objektive für größere Sensoren sind zwar abwärtskompatibel, d.h. sie können auch an Kameras mit kleineren Sensoren verwendet werden. Hier ist allerdings zu beachten, dass sich der Bildwinkel und somit der Bildausschnitt des Motivs ändert. Der sogenannte Cropfaktor entspricht bei 2/3“ zu 1/2“ Sensoren 1,33. Bei 2/3“ zu 1/3“ beträgt er 2. Beim Verhältnis von 1/2“ zu 1/3“ beläuft er sich auf 1,5. Im Bezug auf das 35mm Kleinbildformat wäre der Faktor beim 2/3“ Sensor 2,5, bei 1/2“ wäre er 3,5 und bei 1/3“ entspricht er 5.

So würde beispielsweise ein 16mm Objektiv (Normalbrennweite für 2/3“) an einer 2/3“ Kamera einen Bildwinkel von 55° erreichen. Dasselbe Objektiv an einer 1/2“ Kamera würde nur einen Winkel von 43° und noch einmal das selbe Objektiv an einer

1/3“ Kamera einen Bildwinkel von 29,5° erreichen. Je kleiner also der Sensor wird um so kleiner wird auch der abgebildete Bildausschnitt. Die für ein System in Frage kommenden Kameras haben überwiegend 2/3“ oder 1/3“ Sensor. Die Kamera von Ikonoskop hat einen ca. 1/2“ großen Sensor, was in etwa dem analogen 16mm Filmformat entspricht.

4.2.3 Seitenverhältnis und Sensor

Während der Recherche ist ebenfalls aufgefallen, dass einige Objektive ausschließlich für eine Abbildung im 16:9 Format produziert werden und andere sich nur auf das 4:3 Format beschränken. Da die gängigen HD-Standards allerdings für das 16:9 Seitenverhältnis ausgelegt sind, wären letztere Objektive nicht für ein solches System geeignet. Die Abmessungen des Sensors geben hier die erforderliche Beschaffenheit vor. Die Qualität des Bildes kann ebenfalls nicht gewährleistet werden, wenn ein Objektiv für Modelle mit nur einem CCD an einer 3-Chip-Kamera verwendet wird. Die Konstruktion der Linsen folgt hier unterschiedlichen Ansprüchen, wie z.B. der Verwendung von Prismen. Um eine geringe Farb aberration zu erreichen sind die Objektive speziell gefertigt. Die Kamerahersteller empfehlen in diesem Fall ausdrücklich Objektive für 3CCD Kameras zu verwenden.

4.2.4 Brennweitenbereich

Für Stereoskopie werden Brennweiten im unteren bis mittleren Bereich, also von leichtem Weitwinkel bis knapp über Normal-Brennweite, benötigt. Um den Effekt des Cardboarding zu vermeiden sollten keine Brennweiten über 50mm für einen 2/3“ Sensor verwendet werden. Dies ergaben die bisherigen Versuche von Eric Schumacher während seiner Arbeit im Sommersemester 2008. Die minimale Brennweite hängt von der Beschaffenheit der Linse selbst ab. Kurze Brennweiten neigen dazu das Bild tonnenförmig zu verzerren. Dies ist eine Folge der leicht gewölbten Linse zum Rand des Objektivs hin. Dadurch kann vertikale Parallaxe zwischen den Stereohalbbildern auftreten und sollte deshalb vermieden werden. Asphärische Objektive könnten hier Abhilfe schaffen, sind aber teurer in der Anschaffung. Welche genauen Brennweiten noch einsetzbar sind, hängt speziell vom verwendeten Objektiv ab und muss in der Praxis getestet werden. Die Firma Easylook System verwendet z.B. Objektive ab etwa 8mm an zwei 2/3“ Kameras.

Für den Einsatz von Prime-Objektiven sollten mindestens 3 verschiedene Brennweiten vorhanden sein, um technisch und gestalterisch genügend Spielraum zu erhalten. Die Festbrennweiten sollten, wie erwähnt, einem leichten Weitwinkel, einer Normalbrennweite und einem leichten Teleobjektiv entsprechen. Die Normalbrennweite hängt vom Format des Kamerasensors ab und entspricht etwa folgenden Werten:

2/3“ = 16mm, 1/2“ = 12mm, 1/3“ = 8mm

4.2.5 Bauart

Objektive werden in verschiedenen Bauarten produziert. Zu diesen zählen Festbrennweiten-, Varifokal- und Zoom-Objektive. Die allgemein bekannten Fest- und Zoom-Bauweisen bedürfen an dieser Stelle keiner Erklärung, lediglich das Varifokal-Objektiv ist weniger geläufig. Hier handelt es sich um Objektive deren Brennweite ebenfalls veränderbar ist. Allerdings muss nach diesem Vorgang, im Gegensatz zum Zoom-Objektiv jedesmal die Schärfe neu eingestellt werden. Diese Bauform ist zwar einfacher und günstiger herzustellen, was sich auch im Endpreis bemerkbar macht, geht aber mit gestalterischen Einschränkungen einher, da Zoomfahrten ausgeschlossen sind. Aus der Recherche zu Objektiven ging hervor, dass im High-Definition-Bereich sehr wenige Zooms existieren, welche an die favorisierten Kameras angeschlossen werden können. Hier stellt der C-Mount eine deutliche Einschränkung dar. Allerdings könnte sich diese Situation in Zukunft verbessern, da die Entwicklung der stereoskopischen Systeme erst in den Anfängen steht. Auch die verwendbaren Kameras sind noch nicht lange auf dem Markt erhältlich. Allerdings lässt sich weiter feststellen, dass die Zoom-Bauweise am kostenintensivsten ist, da an sie hohe mechanische und optische Ansprüche gestellt werden. Festbrennweiten hingegen sind die günstigste Lösung der Objektive, sieht man von den professionellen Cine-Linsen ab. Diese Bauform ist sehr umfangreich in HD-Auflösung erhältlich und liegt für C-Mounts preislich deutlich unter den Zoom-Objektiven. Festbrennweiten stellen außerdem die zuverlässigste Lösung für den Stereobetrieb dar, weil bei ihnen die produktionsbedingten Qualitätsunterschiede am geringsten ausfallen. Sie bringen allerdings den Nachteil mit sich, dass sie während der Dreharbeiten gewechselt werden müssen. Dieser Umstand bringt im Hochschulbetrieb gleich mehrere Probleme mit sich. Zum einen können Abnutzungserscheinungen am Anschluss des Objektivs und der Kamera auftreten. Weiter können durch diese Abnutzung Verunreinigungen auf dem Sensor verursacht werden. Das Risiko der Verschmutzung oder im schlimmsten Fall Beschädigung der Linse oder des Sensors ist durch den Wechselvorgang grundsätzlich höher. Zum anderen haben die Objektive höchstwahrscheinlich unterschiedliche Abmessungen, was die mechanische Einstellung über das Rig deutlich erschwert oder sogar unmöglich machen könnte. Daher könnten Varifokal-Objektive womöglich doch die bessere Wahl für den Lehrbetrieb sein.

4.2.6 Bedienung

Bei der Steuerung der Objektivparameter Blende, Zoom und Fokus werden drei unterschiedliche Typen von Objektiven angeboten. Elektronische Objektive werden über eine separate Steuerung bedient. In den allermeisten Fällen findet dies bei Consumer-Kameras Anwendung, aber auch im Professionellen Bereich werden elektronische Steuerungen verwendet. Der Unterschied zwischen den beiden Bereichen ist jedoch sehr groß. Schulterkameras verfügen so gut wie immer über Servosteuerungen am Objektiv, um die Parameter für die Brennweite zu verändern. Externe Steuerelemente machen auf diese Weise Objektive auch aus größerer Entfernung bedienbar. Durch die Elektronik wird das Objektiv zwar teurer, kann aber als späteres Upgrade des Systems

gesehen werden, falls von der mechanischen Steuerung abgewichen werden soll. Automatische Objektive reagieren auf die Umgebungsverhältnisse und passen sich ihnen bestmöglich an. Es liegt auf der Hand, dass diese Art der Bedienung für eine stereoskopische Filmproduktion nicht geeignet ist. Der Kameramann hat keinerlei Einfluss auf die Parameter und kann so auch nicht die Synchronität der Objektive gewährleisten.

Als letzte Variante sind die manuellen Objektive zu nennen, welche als einzige Bedienungsart für das Stereosystem einsatzfähig sind. Alle Parameter werden von Hand verändert. Für die Auswahl der Objektive musste festgestellt werden, dass es bisher nur wenige manuelle Zoomobjektive gibt, Festbrennweiten hingegen überwiegend manuell bedient werden. Bei den Zooms findet meist die elektronische Steuerung Anwendung.

4.2.7 Qualität

Im Bereich der C-Mount Kameras werden bisher auch wenige Objektive in hoher Auflösung gefertigt. Dies kann wiederum daran liegen, dass die HD-Technik noch nicht überall eingesetzt wird und auch daran, dass erst in der letzten Zeit verstärkt hochauflösende Kameras in dieser Sparte entwickelt wurden. Die Objektive werden statt mit MTF-Messkurven teilweise mit Megapixelangaben oder mit der unbestimmten Bezeichnung High Quality beschrieben. Diese Angaben sind oft nur unbefriedigend und haben bisher wenig Bezug zu den existierenden Standard-HD-Formaten. Wie auch bei anderen Faktoren erhöht sich der Preis bei hochauflösenden Objektiven enorm.

Im professionellen Betrieb werden die Objektive für stereoskopische Verwendung zusätzlich vermessen und die Pärchen mit eindeutigen IDs versehen. Damit wird sichergestellt, dass beide die exakt gleichen optischen Eigenschaften besitzen. Spätere Unterschiede der aufgezeichneten Videos für linkes und rechtes Auge werden dadurch vermeiden. Auch die Verzerrung am Rand der Objektive sollte möglichst gering ausfallen, um die schon erwähnten Tonnenverzerrungen gering zu halten. Dies lässt im Betrieb die Verwendung von kürzeren Brennweiten zu.

4.2.8 Abmessungen

Die Abmessungen der Objektive haben bei einem Side by Side Rig direkten Einfluss auf das Stereosystem. Ist der größte Durchmesser über 65mm, kann die angestrebte Stereobasis nicht erreicht werden. Im Fall der mechanischen Steuerung und dafür angepasster Zahnräder müssen diese in die Berechnung mit einbezogen werden. Falls das Objektiv schon mit Zahnrädern ausgestattet ist, entfällt dies natürlich. Als groben Maximal-Richtwert für nicht verzahnte Gehäuse kann man 50mm Durchmesser angeben. Mit diesem Wert bleibt genug Distanz für Zahnräder und Lauffreiheit, während gleichzeitig die Stereobasis eingehalten wird.

4.2.9 Preis und Objektivmodelle (2009)

Der Preis für ein Objektiv hängt von fast allen oben genannten Faktoren und deren Kombination ab und bewegt sich bei HD-Formaten etwa zwischen 300€ bis über 10000€. Im SD-Bereich sind Objektive teilweise billiger als 300€. Eigenschaften wie Zoom und Lichtstärke fallen hier stark ins Gewicht, wohingegen die Auswahl des Anschlusses wenige Auswirkungen zeigt.

Es wurden mögliche Objektive für das System ausgesucht, deren Anschaffung aus Kostengründen und technischen Anforderungen erst sinnvoll sind, wenn die Kameras des Systems festgelegt wurden.

Zoombauweise:

- Fujinon - T16X5.5DA-R11 (ca. 2760€, 1/3“, 5.5-88mm, 3CCD)
- Fujinon - H16X10A-M41 (ca. 1500\$, 2/3“, 10-160mm, 3CCD)
- Pentax - H6Z810 (ca. 580€, 1/2“, 8-48mm, CCD)

Festbrennweite:

- Computar - M1214-MP (ca. 220€, 2/3“, 12mm, CCD)
- Kowa - LM12JCM (ca. 150€, 2/3“, 12mm, CCD)
- Schneider Kreuznach - CINEGON 2.1/6 (ca. 1415€, 2/3“, 6mm, CCD)
- Pentax - C1614-M (ca. 160€, 2/3“, 16mm, CCD)

4.3 Kamera- und Objektivsteuerung

Wie bereits festgestellt ist die Synchronität ein entscheidender Aspekt eines 3D-Systems. Während der Aufnahme wird an den grundlegenden Einstellungen in der Regel nichts verändert, weshalb sie bei jeder Kamera von Hand vorgenommen werden können. Eine Ausnahme stellen die Objektivparameter dar. Um sie zu kontrollieren gibt es die Möglichkeit beide Objektive mechanisch oder elektronisch miteinander zu verbinden. Letztere Methode ist zwar technisch durchführbar, allerdings fehlt oft das Gefühl für die Bedienung. Davon abgesehen ist die elektronische Lösung auch sehr kostspielig und benötigt zusätzlichen Platz an den Objektiven. Systeme wie das FI+Z von Preston oder das Wireless Remote System von Arri könnten zwar grundsätzlich für die Steuerung eingesetzt werden, würden aber den preislichen Rahmen einer Hochschule deutlich überschreiten (ca. 10000€). Die Aufbauten der Steuerungen setzen sich aus einer Bedieneinheit, einem Empfänger und den dazugehörigen Motoren und Akkus zusammen. Voraussetzung für die Funktion sind allerdings Objektive mit integriertem Zahnrad, in die solche Motoren eingreifen können. Zusätzlich verfügen derartige Bedienelemente meist über eine Anzeige der eingestellten Parameter. Um eine solche Kontrolleinheit selbst zu konstruieren wird fundiertes Wissen auf dem Gebiet der Elektrotechnik und Mechanik benötigt. Eine rein mechanische Vorrichtung kann hingegen eigenständig entwickelt werden.



Abbildung 3 – Objektivsteuerungselemente
Preston Cinema Systems HU3, Arri Lens Control, Arri WRS

4.4 Synchronisation

4.4.1 Regieklappe

Die einfachste und älteste Methode zwei Geräte zu synchronisieren stellt die Regieklappe dar. Mit ihr werden üblicherweise Bild und Ton oder mehrere Kameras miteinander synchronisiert. Die Klappe wird so gehalten, dass alle notwendigen Geräte sie im Bild haben. Dann wird die Klappe, für die Synchronisation des Tons und Bilds, zugeschlagen. In der Nachbearbeitung werden die einzelnen Clips dann an dieser Stelle identisch ausgerichtet. Professionelle Varianten von Klappen können beispielsweise auch den Timecode anzeigen. Mit ihnen gekoppelte Kameras können so den TC an die Klappe weitergeben. Dadurch kann im Schnittprogramm zusätzlich der TC als Referenz für mehrere Kameras dienen. Auch in den bisherigen Stereoproduktionen der Hochschule kam die Regieklappe schon zum Einsatz. Zur Sicherheit kann die Regieklappe auch zusätzlich zu anderen Verfahren verwendet werden, um eventuellen Problemen vorzubeugen.

4.4.2 LANC

Das Local Application Control Bus System oder kurz LANC ist eine bidirektionale Schnittstelle der Firma Sony. Eingeführt wurde sie um zwei Geräte zu synchronisieren und fernzusteuern. Der Umfang der Befehle umfasst unter anderem die Steuerung der Brennweite, der Aufnahme und das Auslesen des Timecodes. Die Klinkenbuchse wird in Kameras der Hersteller Sony, Canon und JVC eingesetzt und hauptsächlich im Consumer- und Prosumerbereich verbaut. Auch bei einigen Fotokameras werden LANC-Anschlüsse verwendet. Von Amateuren wird diese Art der Synchronisation im Stereobereich oft verwendet, da sie eine zuverlässige, relativ genaue und meist kostengünstige Kontrolle zweier Kameras zulässt. Es werden auch spezielle Bedienteile für die 3D-Produktion angeboten, mit denen zwei Kameras parallel angesteuert werden können.

4.4.3 Direct Show

Wie in der Arbeit von Volker Stephan bereits beschrieben, kann eine Anwendung in Direct Show ebenfalls zur Synchronisation von Kameras eingesetzt werden. Die Kameras werden über HD-SDI Verbindungen an die Blackmagic Decklink Videokarten und somit an den Aufzeichnungsrechner angeschlossen. Dort wird der Videodatenstrom dann von einem Filtergraphen verarbeitet. Im Graphen ist ein Capturefilter der Videokarten enthalten, sowie ein AVI-Muxer und ein Filewriter. Um jetzt die aufzunehmenden Videos zu synchronisieren, wird zusätzlich ein Audiocapture-Filter mit einer Reference-Clock hinzugenommen. Das Synchronisationssignal der Reference-Clock sorgt nun für einen identischen Takt der beiden Aufnahmestränge.¹⁶ In der Praxis funktioniert das Verfahren sehr gut und zuverlässig. Allerdings ist der Aufbau auf einen Computer angewiesen, der die geforderte Leistung zur Aufnahme erreicht.

4.4.4 Genlock

Genlock wird bei professionellen Kameras für die Synchronisation mehrerer Geräte eingesetzt. Dabei werden alle angeschlossenen Geräte auf einen gemeinsamen Takt synchronisiert, der intern oder separat generiert wird. Durch diese Einrichtung können die Bilder bei einer Multikameraproduktion miteinander vermischt werden, ohne einen zeitlichen Versatz der Zeilen oder gar der Bilder zu riskieren. Je nach Kameraformat wird ein anderes Sync-Signal benötigt. SD-Systeme werden über ein Bi-Level-Signal, dem sogenannten Blackburst gekoppelt. Im HD-Bereich wird ein Tri-Level-Signal verwendet. Viele Kameras die in der Recherche als relevant eingestuft wurden, haben die Möglichkeit Genlocking zu verwenden. Das Signal wird über eine spezielle BNC-Schnittstelle an die Kamera übertragen. Geräte wie der SPG 1701 von LYNX Technik oder der GEN10 von AJA (Abb. 6) wären für den Einsatz in einem mobilen, stereoskopischen System einsetzbar. Allerdings ist zu beachten, dass die Funktion des Genlockings auch von den verwendeten Videoformaten abhängt. Momentan erhältliche Geräte liefern aber in der Regel alle gängigen Standards. Sync-Boxen sind von einer eigenen Stromversorgung abhängig, die meist extern über ein Netzteil realisiert wird.



Abbildung 4 – Genlockgeneratoren
LynxTechnik Yellowbrick und AJA GEN10

¹⁶ Vgl. Volker Stephan 2008, S. 50–51

4.4.5 Timecode

Für die Synchronisation zweier Kameras kann ebenso der erzeugte Timecode verwendet werden. Die meisten Kameras liefern bei der Übertragung des Videosignals einen Zeitstempel, und somit eine genaue Referenz für jedes aufgenommene Bild. Falls die Synchronisation per Timecode vorgenommen werden soll, wird ein Gerät als Master eingerichtet, welches dann den Referenz-Timecode liefert, und ein Gerät als Slave, das sich nach dem empfangenen TC richtet. Jedoch sind nicht alle Kameras in der Lage Timecode abzusetzen. Er wird von einem Recorder generiert, wie er in den meisten Kameras eingebaut ist. In den Mikro-Kameras, die für Stereoskopie verwendet werden müssten, ist jedoch selten eine Aufnahmemöglichkeit integriert. Um dennoch eine Synchronisation über den Timecode zu realisieren kann ein TC-Generator eingesetzt werden, der das Signal wiederum in den Rekorder einspeist und so synchronisiert. Auf diese Weise könnten auch zwei Aufnahmegeräte synchronisiert werden. Das würde den Vorteil mit sich bringen, dass ein geringerer Datendurchsatz pro Gerät gehandhabt werden muss. Wie auch bei den Genlock-Generatoren ist für den Betrieb eines Timecode-Generator eine separate Spannungsquelle notwendig. Die an den Kameras vorgesehenen Framerates müssen ebenfalls vom TC-Generator unterstützt werden, um fehlerfrei zu synchronisieren.

4.5 Kameramontage

Um zwei Kameras für stereoskopischen Film zu verwenden und einzustellen werden sogenannte Stereo-Rigs verwendet. Diese Rigs sind nötig um die Kameras auf ein Stativ zu montieren. Hinzu kommt dass sie die Stereobasis der Kameras variabel halten und gleichzeitig das System im Gleichgewicht bleibt. Da die stereoskopische Produktion noch in den Anfängen steckt, existieren bisher nur wenige kommerzielle Produkte auf dem Markt. Neben diesen gibt es noch einige Aufbauten von engagierten Amateuren.

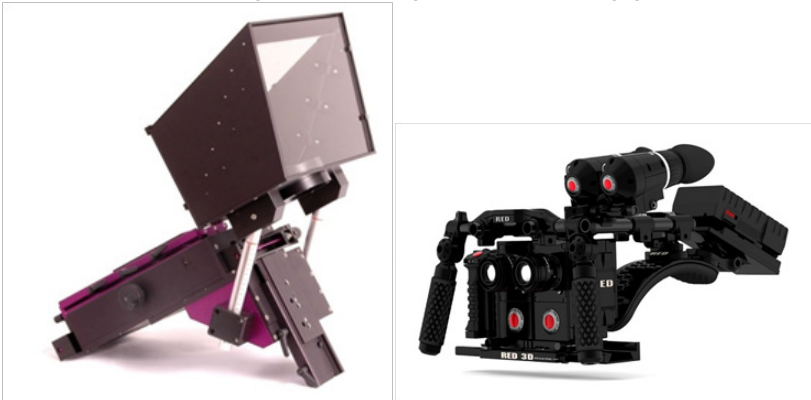


Abbildung 5 – Stereorigs
Swissrig und RED

4.5.1 Variationen

Es existieren zwei verschiedene Arten von Rigs. Eine mögliche Variante ist ein Mirror-Rig. Sie sind so konstruiert, dass die Kameras im 90° Winkel zueinander angeordnet sind. Das hat den Vorteil, dass die Größe der Kameragehäuse und Objektive keinen Einfluss auf die einstellbare Stereobasis hat. Es können alle Film- oder Videokameras montiert werden, die auch im normalen Filmbetrieb eingesetzt werden. Das Bild wird im Fall eines Mirror-Rigs über einen Strahlenteiler oder einen Spiegel an die beiden Kameras verteilt. Dazu ist der Spiegel im 45° Winkel zu beiden Objektiven angebracht. Die Kameras können in diesen Vorrichtungen jede Stereobasis bis zur völligen Überdeckung erreichen. Eine solche Ausführung eines Rigs ist je nach Kamera relativ groß, was auch das Gewicht und die mechanischen Torsionskräfte des Aufbaus beeinflusst. Die Möglichkeit einer manuellen Objektivsteuerung ist durch die Bauweise sehr erschwert.

Eine weitere Bauform ist die Side by Side Variante, welche wie der Name sagt eine Kameraanordnung nebeneinander vorsieht. Hier wird davon ausgegangen, dass die Kameras klein genug sind, um eine Stereobasis von mindestens 6,5cm zu erreichen. Die Objektive sind in diesem Fall parallel angeordnet, können allerdings auch konvergent ausgerichtet werden. Dies setzt eine spezielle Konstruktion des Rigs voraus. Bei der Anordnung nebeneinander muss darauf geachtet werden, dass die Kameras exakt gleich hoch und im selben Winkel angebracht sind. Abweichungen können später den räumlichen Eindruck zerstören. Das gilt auch für die Ausrichtung der Spiegelvariante. Die Recherche zu Rigs ergab, dass auch elektronisch verstellbare Rigs in dieser Bauweise produziert werden. Sie können die Stereobasis motorisiert während der Aufnahme verändern. Ein anderer Ansatz ist die motorisierte Verstellung der Konvergenz, wie er von James Cameron angewandt wird.

4.5.2 Vorausgehende Erfahrungen

Bisher wurde die schon zuvor erwähnte Stereoschiene für die Montage der Kameras verwendet. Aus den entstandenen Arbeiten und durchgeführten Befragungen geht hervor, dass die parallele Ausrichtung der Kameras und die Einstellung der Stereobasis nicht ganz einfach von statten gingen. Die Parallelität wurde mit Hilfe einer geraden Latte oder Winkels ausgerichtet. Der Gegenstand musste flach vor die Objektive oder im Fall des Winkels an die Stativplatte gehalten werden erst dann wurde die Kamera positioniert. Die Stereobasis ermittelte man mit Hilfe eines Lineals oder Meterstabs und stellte sie ein. Auf der Stereoschiene waren Stativplatten befestigt, in die dann die Kameras eingerastet wurden. Für den Fall, dass die Stativplatten sich verschoben, musste alles neu eingestellt werden.

Das bisher verwendete System war durch die eingesetzten JVC Kameras auch sehr unhandlich und schwer. Die Baugröße der Kameras beschränkte die minimale Stereobasis auf 13,8 cm wobei schon Komponenten wie Sucher und Mikrofonhalterung der

Geräte demontiert werden mussten.¹⁷

4.5.3 Anforderungen

4.5.3.1 Material und Bauweise

Um ein System für die Hochschule zu konstruieren müssen einige Faktoren bedacht werden. Das Rig soll später auch im Handkamerabetrieb und nicht nur auf einem Stativ verwendet werden können. Das heißt es muss klein, leicht und einfach zu bedienen sein. Da an der Hochschule voraussichtlich nur ein einziges Stereorig vorhanden ist, muss es für alle Situationen bestmöglich geplant werden. Es soll also sowohl in kleinen Räumen als auch freier Natur eingesetzt werden können. Hier muss also Korrosion, Langlebigkeit und Baugröße beachtet werden. Auch die Balance des Aufbaus für den Handkamerabetrieb sollte bedacht werden, damit die Bedienung möglichst angenehm ist.

4.5.3.2 Einstellungen

Die vorherigen Probleme sollen weitestgehend beseitigt und zusätzlich eine Vorrichtung für die synchrone Einstellung der Kameraobjektive realisiert werden. Diese soll den gestalterischen Spielraum erweitern und Schärfenverlagerungen oder Zooms ermöglichen. Durch das Erreichen von kleineren Stereobasen wird auch der Einsatz von Konvergenzen interessant, weshalb die Möglichkeit von Winkelstellungen der Kameras integriert wird. Alle Parameter sollen möglichst leicht und ohne Werkzeug verändert werden können, so dass keine langen Einstellungsphasen während der Dreharbeiten entstehen.

4.6 Datentransfer

Für die Aufnahme von stereoskopischem Video ist in vielen Fällen eine Übertragung des Signals von den Kameras zu einem Aufzeichnungsgerät notwendig, da die Videodaten nicht mehr intern in der Kamera gespeichert werden können. Dies hängt mit der Kompression und der Größe der Bilder zusammen. Die entstehenden Datenvolumina benötigen auch entsprechende Übertragungswege. Hier gibt es verschiedene Möglichkeiten.

4.6.1 Firewire

Der von Apple eingeführte Standard IEEE 1394 ist eine Möglichkeit Daten schnell zwischen einer Kamera oder anderen externen Gerät und einem Computer zu übertragen.¹⁸ Eingesetzt wird die Schnittstelle an verschiedenen Arten von Kameras wie DV-Camcordern, Webcams und Industriekameras. An fast allen neuen Mac-Rechnern werden mittlerweile sogar die S800 Schnittstellen eingebaut. Verschiedene Ausprägungen mit unterschiedlichen Bandbreiten bedienen die speziellen Anforderungen der Geräte. Der Standard IEEE 1394b (S800) lässt derzeit die höchste Datenrate mit

¹⁷ Vgl. Volker Stephan 2008, Eric Schumacher 2008, Anja Frank 2007

¹⁸ Vgl. Hardware & Drivers - FireWire

bis zu 800Mbit pro Sekunde zu. Auch eine Erweiterung auf 3,2Gbit soll eingeführt werden. Die besonderen neunpoligen Steckverbindungen und Kabel sollen identisch mit der 1394b Schnittstelle sein, allerdings existieren für die S3200 Verbindung noch kaum Geräte die dies unterstützen.¹⁹ S800 wird jedoch bei einigen relevanten Industriekameras verwendet und muss deshalb in dieser Arbeit berücksichtigt werden. Um zwei Kameras parallel an einem Computer anzuschließen ist sicherzustellen, dass jede Kamera über einen eigenen Bus betrieben wird. Anderenfalls könnte die Bandbreite reduziert und so die Performanz beeinträchtigt werden. Um das zu erreichen, müssen zwei separate Controller-Karten oder eine Karte mit zwei Bussen verwendet werden. Die Kabellängen von bis zu 100m sind mehr als ausreichend und können durch die Verwendung von Lichtleitern noch einmal erhöht werden. Von Vorteil könnte die interne Stromversorgung bei Firewire-Geräten sein. Es sind keine zusätzlichen Akkus oder Netzteile notwendig um die Kameras innerhalb der 100 Meter Entfernung zu betreiben.

4.6.2 Ethernet

Auch die Netzwerktechnik bietet in speziellen Ausführungen eine adäquate Übertragungsrate. Das Gigabit-Ethernet hat dem Namen entsprechend eine Bandbreite von 1Gbit pro Sekunde und wird in Kupfer- oder Glasfaserausführung hergestellt. Die beiden Varianten unterscheiden sich aber lediglich in ihrer möglichen Kabellänge, die bei Kupfer 100m und bei Glasfaser zwischen 200m und 550m liegt. Der zweite relevante Standard ist das 10GE, welcher eine Übertragungsrate von 10Gbit pro Sekunde erreicht. Diese Verbindungen reichen aus, um einen HD-Videodatenstrom zu übertragen und werden in der Praxis auch von verschiedenen Überwachungs-, Industrie und Profikameras eingesetzt. Der benötigte IP-Overhead ist dabei kein einschränkender Faktor mehr. Ein weiterer Aspekt ist Power over Ethernet. Ähnlich wie bei der Firewire-Schnittstelle kann auch hier eine Stromversorgung direkt über die Netzwerkleitung eingespeist werden. Unter dem Standard IEEE 802.3af wird für angeschlossene Geräte eine Spannung von 48 Volt und bis zu 12,95 Watt bereitgestellt. Die maximale Leistungsaufnahme der angeschlossenen Geräte darf im Dauerbetrieb 350 Milliampere nicht überschreiten. Für den Betrieb einer PoE-Lösung sind spezielle Switches notwendig, die mit entsprechenden Line Cards zur Spannungsversorgung ausgestattet sind.²⁰ Außerdem muss das zu betreibende Gerät, z.B. eine Netzwerkkamera, ebenfalls speziell auf diese Betriebsart ausgerichtet sein.

4.6.3 HD-SDI

Die Datenleitung HD-SDI nach der Norm SMPTE 292M ist die momentan am weitesten verbreitete Verbindung zur Übertragung von Videosignalen im professionellen High Definition Bereich. Über die Schnittstelle sind Bild und Ton sowie Metadaten in Form von Timecode oder Text übermittelbar. Drei unterschiedliche Ausprägungen des Serial Digital Interface werden produziert. Diese sind unter den Namen HD-SDI

¹⁹ Vgl. Windeck 17.12.07

²⁰ Vgl. Katharina Hörnstein 22. Januar 2005

(SMPTE 292M) mit einer Datenrate von 1,485Gbit/s, Dual Link HD-SDI (SMPTE 372M) mit einer Datenrate von 2,97Gbit/s und 3G-SDI (SMPTE 424M) ebenfalls mit einer Datenrate von 2,97Gbit/s. Bei Dual Link werden zwei HD-SDI Verbindungen gekoppelt. Dies ermöglicht eine höhere Übertragungsrate. Die beiden übrigen Standards kommen hingegen mit jeweils nur einem Kabel aus. Die Erweiterung des HD-SDI auf Dual Link bringt den weiteren Vorteil, dass ein breiteres Spektrum an Videoformaten unterstützt wird. So können mit Duallink und 3G-SDI auch 1080p und 2k Formate übertragen werden. Bei einer Standardverbindung kann maximal ein YCbCr 4:2:2 Signal mit 10 Bit übertragen werden. Duallink und 3G-SDI Signale ermöglichen hingegen unkomprimierte 10bit oder 12bit YCbCr und RGB Formate. Die Anschlüsse sind als BNC-Buchsen gefertigt, die auch schon bei der SDI Variante eingesetzt wurden. Heutzutage gibt es sehr viele Kameras und Videogeräte, welche über HD-SDI Ein- und Ausgänge verfügen.

4.6.4 HDMI

Der für den Heimkinobereich entwickelte Standard HDMI, als digitale Schnittstelle, bietet wie auch die vorherigen Übertragungswege eine hohe Datenrate. HDMI wird in vielen verschiedenen Ausführungen hergestellt. Die neueste Weiterentwicklung, Version 1.4, wurde im Mai 2009 vorgestellt und im Juni 2009 eingeführt.²¹ Mit dieser Version können Datenraten von bis zu 8,16Gbit/s und einem 4k Bildformat erreicht werden. Die älteren Ausführungen übertragen nur kleinere aber immer noch ausreichende Videoformate b.z.w. Datenraten von 3,96Gbit. Kabellängen von 15 Metern (Fiberglas bis 100m) sind allerdings nicht besonders lang und bieten für ein Kabelgebundenes System wenig Spielraum. Vor allem im Consumerbereich, aber auch bei einigen semiprofessionellen Kameras werden HDMI-Schnittstellen eingesetzt. Der integrierte HDCP Kopierschutz der Verbindung stellt allerdings ein Hindernis für die Aufzeichnung dar. Dieser kann in der Regel nur von Wiedergabegeräten und nicht von Aufzeichnungsgeräten dekodiert werden. Bei der neusten Version von HDMI wird sogar eine Übertragung von 3D in den unterschiedlichsten Formaten wie z.B. Side by Side und 2D mit Tiefeninformation unterstützt. Diese Anwendung bleibt allerdings Geräten vorbehalten, welche als Quelle für räumliche Inhalte oder zur Wiedergabe dieser bestimmt sind.²²

4.7 Aufnahme

Die Aufzeichnung stereoskopischer, unkomprimierter Videos ist momentan noch nicht weit verbreitet. Aus diesem Grund werden auch nur wenige Lösungen zur Handhabung der sehr hohen Datenmengen angeboten. Viele dieser Lösungen sind nur für den Studiobetrieb ausgelegt und können nicht im mobilen Einsatz verwendet werden. Oft arbeiten Rekorder mit einem internen Codiervorgang und nehmen kei-

²¹ Vgl. Steve Venuti 2009

²² Vgl. HDMI Licensing 22.07.2009

ne unkomprimierten Daten auf. Nur wenige Systeme erreichen überhaupt die nötige Aufzeichnungsgeschwindigkeit und Speicherkapazität um den Datenstrom zweier HD-Kameras zu verarbeiten. Es muss mit Datenraten von 46MB/s (720p25, 4:2:2, 8bit) bis 260MB/s (1080p50, 4:2:2, 10bit) gerechnet werden, die bei HD- Video pro Kamera auftreten können.

4.7.1 RAID

Eine Möglichkeit den Videodatenstrom aufzuzeichnen besteht in einem Raid-0 Verbund, also der Kopplung mehrerer Festplatten eines Computers, zur Erhöhung der Lese- und Schreibgeschwindigkeit. Die Daten werden auf die vorhandenen Festplatten verteilt und dadurch die Schreib- und Lesegeschwindigkeit erhöht. Die Schreibgeschwindigkeit kann dabei für zwei Festplatten maximal doppelt so hoch sein wie im Einzelbetrieb. Bei mehreren Festplatten erhöht sich der Wert entsprechend. Die Geschwindigkeit hängt aber unter anderem von den zu schreibenden Daten und der Zugriffszeit der Festplatte ab. Außerdem wirken auch andere Faktoren auf den Raid-Verbund ein, die sich begrenzend auf die Datenrate auswirken können. Darunter wäre natürlich der verwendete Raid-Controller zu beachten, der je nach Leistung und Bauweise die Anzahl der anschließbaren Geräte und die Datenübertragungsrate begrenzt. Auf dem Markt sind Controller mit bis etwa 600MB/s Durchsatz erhältlich.

Da die meisten Raid-Controller über einen PCIe Steckplatz betrieben werden ist auch dieser eine mögliche Engstelle des Verbundes. Die PCIe-Schnittstelle erreicht ca. 250MB/s. PCIe 2.0 kann Datendurchsätze von ca. 500MB/s verarbeiten. Dies kann aber durch die Kopplung mehrerer Lanes eines Steckplatzes beschleunigt werden, wobei sich bei Verdoppelung der Lanes auch die Übertragungsrate verdoppelt. Auf diesem Weg können schon für PCIe 1.0 bis zu 8GB/s erzielt werden.²³ Der Puffer für die Übertragung ist also ausreichend.

Die Anschlüsse und Verbindungen der Festplatte sind in diesem Kontext ebenfalls zu berücksichtigen. Die mittlerweile am häufigsten verwendete Schnittstelle Serial ATA bietet aber eine sehr hohe Bandbreite von 300 bis zukünftig 600MB/s, so dass sich auch dies nicht begrenzend auswirkt.²⁴ IDE Festplatten sind etwas langsamer, da hier die Übertragungsrate nur bei etwa 133MB/s liegt. Neben diesen Verbindungen gibt es noch weitere, die z.B. in Servern Verwendung finden. Da diese nur in der entsprechenden Umgebung angewendet werden und deshalb kein mobiles System unterstützen sollen sie hier nicht genauer beschrieben werden.

Der letzte und schwerwiegendste Engpass für Daten entsteht in der Festplatte selbst. Heutige Laufwerke sind durch die mechanische Einschränkung der Geschwindigkeit auch in ihrer Schreibgeschwindigkeit begrenzt. Dies macht erst die Installation eines Raid-Verbundes notwendig. Durchschnittliche Datenraten von etwa 120 bis 150MB/s

23 Vgl. PCI-Express – Wikipedia, 16.07.2009

24 Vgl. The Serial ATA International Organization 24.07.2009

werden derzeit für einzelne Festplatten im Schreibvorgang erreicht. Durch den Einsatz verschiedener Zwischenspeicher, dem Cache, kann die Leistung jedoch variieren. Insgesamt stellen Festplatte und RAID-Controller die wahrscheinlich kritischen Komponenten eines solchen Systems dar. Unabhängig von der Speicherleistung des Systems erhöht sich bei einem Raid-0-Verbund auch immer die Ausfallgefahr der Festplatte. Je mehr Laufwerke in einem Verbund geschaltet sind desto höher die Wahrscheinlichkeit eines Ausfalls. In einem solchen Fall würde dies den Verlust aller gespeicherten Daten bedeuten, da sie auf alle Festplatten verteilt aufgezeichnet werden.²⁵ Um ein solches System transportabel zu halten, müsste es in ein Notebook integriert werden. Notebooks mit kleinen Festplattenraids aus 2 Discs werden zwar angeboten, größere interne Verbünde sind jedoch aus Platzgründen schwierig.

4.7.2 Solid State Festplatten

Eine weitere Art der Speicherung stellen Solid-State-Drives dar. Diese Flash- oder SDRAM-Speicher benötigen keine beweglichen Teile und bietet dadurch eine geringe Latenzzeit, die wiederum für einen hohen Datendurchsatz und somit auch eine hohe Schreibgeschwindigkeit sorgt. Aktuelle Werte liegen bei etwa 200MB/s schreibend. Auf dem Markt werden aber beispielsweise von der Firma OCZ auch fertige Raid 0 SSD-Verbünde vertrieben, die eine Lesegeschwindigkeit von 600MB/s und eine Schreibgeschwindigkeit von 500MB/s erreichen und über einen PCI Steckplatz angeschlossen werden. Die Preise entsprechender Geräte sind allerdings noch sehr hoch, was auch für einzelne SSDs zutrifft.²⁶ Die restlichen Anforderungen an Schnittstellen entsprechen denen eines Festplatten Raids.

4.7.3 Rekorder

Die Komplettlösung zur Speicherung von Videodaten stellen Feldrekorder dar. Von unterschiedlichen Herstellern gefertigt bieten sie eine unabhängige Möglichkeit zur Videoaufzeichnung an. Dabei sind die Gehäuse und Anschlüsse schon für den mobilen Einsatz konstruiert. In vielen Fällen sind Funktionen wie die Wiedergabe des aufgenommenen Materials und die Übertragung an einen Rechner ebenfalls integriert.

Vorhandene Anschlüsse sind an die Kamerastandards angepasst und der Betrieb kann mittels optionalen Akkus erfolgen. Intern arbeiten die Rekorder überwiegend mit der schon im Vorhergehenden genannten SSD-Speicherung oder mit herkömmlichen Festplatten. In seltenen Fällen wird auf Band aufgezeichnet, was auch in der 3D-Praxis schon durchgeführt wurde.²⁷ Jedoch gibt es Unterschiede in der Aufnahmeart. Je nach Ausführung wird das aufzunehmende Video komprimiert, um Speicherplatz zu sparen und die notwendige Schreibgeschwindigkeit zu erreichen. In manchen Fällen können nur bestimmte Eingangsformate aufgezeichnet werden. Einige Geräte zur Verarbeitung unkomprimierter Daten sind ebenfalls erhältlich, derzeit aber noch sehr

25 Vgl. RAID – Wikipedia, 23.08.2009

26 Vgl. heise online-Preisvergleich: Solid State Drives (SSD) / Deutschland, 25.07.2009

27 Vgl. Trevor Boyer Feb 18, 2009

teuer. Im Bereich der Stereoskopie, bei welcher zwei Datenströme parallel aufgezeichnet werden müssen, wird die Auswahl noch geringer. Falls Kompressionsverfahren integriert sind, werden überwiegend MPEG und JPEG2000 Codecs eingesetzt.

Bei unkomprimierten Aufnahmen wird in den meisten Fällen 4:2:2 Chroma-Subsampling durchgeführt. Vor allem für den Stereobetrieb ist die Aufnahme von RGB 4:4:4 Material so gut wie nicht möglich, da in diesem Fall zwei Dual Link-Verbindungen verwendet werden müssten. Dies ist bis jetzt in sehr wenigen, mobilen Aufnahmege­rät gewährleistet. Die Anforderungen für das geplante unkomprimierte System sind sehr hoch und lassen wenig Spielraum. Deshalb kommen für den mobilen Einsatz eigentlich nur Feldrekorder in Frage, da sie die geforderte Leistung erreichen. Kriterien, die sich aus den bisherigen Erkenntnissen ableiten lassen, wären:

- Parallele HD Aufnahme zweier HD-SDI Signale
- Synchrone Aufnahme beider Signale
- Unkomprimierte Speicherung
- Akku-Betrieb
- Klein und leicht

Die durchgeführte Recherche hat ergeben, dass nur wenige Geräte geeignet sind.²⁸ Aus einem Telefongespräch mit der Firma Easylook System ging weiter hervor, dass die Aufnahme­geräte nicht generell stereotauglich sind, wenn sie zwei HD-SDI Eingänge besitzen. Diese Funktion ist Softwareabhängig und muss speziell dafür programmiert werden. Die Rekorder von Colorspace (The Icon), DVC (Megacine), Codex Digital (Portable / JPEG2000 Komprimierung), Sony (SRW-1 / Bandgerät) und s.two (OB-1) erfüllen die Kriterien für Stereoskopie grundlegend, liegen jedoch in der höchsten Preisregion bei etwa 20000€ bis 40000€.

In der Praxis werden aus preislichen Gründen auch häufig zwei Flash XDR von Convergent Design verwendet, indem sie durch einen Timecode-Generator synchronisiert werden. Dieses Modell hat jedoch den Nachteil, dass die Videosignale mit einem MPEG2 Codec komprimiert werden, was nicht dem Ziel des Systems entspricht, dafür mit 6000€ pro Stück preislich etwas niedriger angesiedelt ist. Denkbar für einen Aufbau ist an dieser Stelle auch der kleinere, noch in der Entwicklung steckende Nachfolger, Flash Nano. Die Synchronisation zweier Rekorder per Timecode stellt im Allgemeinen eine einfache und zuverlässige Möglichkeit der Kombination zweier Aufnahme­geräte dar. Bei der Verwendung der SRW-1 hingegen, muss darauf geachtet werden, dass die Kameras selbst sehr genau synchronisiert sind.

Die notwendige Stromversorgung kann bei allen Apparaten mit einem Akku hergestellt werden, welcher in manchen Fällen extern und in anderen Fällen im Rekorder

28 Siehe Anhang 3

selbst integriert ist. Passende Batterien sind meist als Zubehör erhältlich. Die Befestigung oder der Transport ist aufgrund der Größe und des Gewichts sehr unterschiedlich. Während der Portable eine Umhängeschlaufe besitzt könnte ein bzw. zwei FlashXDR in einer großen Jackentasche untergebracht werden. Größere Lösungen wie der Icon oder der SRW-1 sollten ihren Platz am besten neben der Kamera oder mit einem langen Verbindungskabel in einiger Entfernung finden. Für einen sehr mobilen Einsatz mit langen Strecken müsste für die größeren Gehäuse eine Art Rucksack verwendet werden.



Abbildung 6 – Feldrekorder
Codex Digital Portable, DVC Megacine, Convergent Design FlashXDR

4.8 Vorschaumöglichkeiten

Um Einstellungen an der Kamera vorzunehmen und den Bildausschnitt oder die Schärfe der Einstellung zu kontrollieren, ist eine Art der Vorschau notwendig. Diese kann nur über zwei unterschiedliche Komponenten realisiert werden und wird im Idealfall, soweit machbar, nach der Aufnahme abgegriffen. Feldrekorder besitzen häufig einen Monitorausgang, der den Anschluss eines Kontrollgeräts ermöglicht. Es muss abgewogen werden inwiefern die Bilder beider Kameras angezeigt werden sollen. Denkbar ist eine Umschaltung zwischen den Vorschaubildern der linken und rechten Kamera oder eine stereoskopische Vorschau der kombinierten Ansichten, wobei das Signal evtl. Aufgespalten werden muss.

4.8.1 Monitor

Eine Grundlegende Entscheidung stellt die Auswahl zwischen stereoskopischem oder herkömmlichem Monitor dar. Für den Fall eines normalen Vorschaumonitors sind einige Voraussetzungen zu erfüllen. Das Gerät sollte neben adäquaten Anschlüssen, meist SDI oder Komponenten-Anschluss, eine umfangreiche Formatauswahl besitzen. Die Monitore erkennen das Eingangssignal in der Regel automatisch und stellen die Wiedergabe entsprechend ein. Allerdings kann nicht jeder Monitor jedes Eingangssignal verarbeiten. Beispielsweise muss ein HD- Format vorher umgewandelt werden um es auf einem SD-Monitor anzeigen zu können. In anderen Fällen muss für die Aufnahme z.B. eine kompatible Framerate gewählt werden. Da hochauflö-

sende Formate im 16:9 Seitenverhältnis arbeiten, sollte auch das Anzeigegerät diese Eigenschaft unterstützen, da andernfalls die Gefahr einer Verzerrung besteht. Ideal könnte sich ein Gerät mit mehreren Anschlüssen und dadurch umschaltbaren Kanälen einfügen, was somit die einfache Kontrolle beider Stereohalbbilder ermöglicht. Für ein mobiles System ist es außerdem wünschenswert, dass der Monitor über eine unabhängige Stromversorgung mittels Batterie bzw. Akku verfügt. Dies wirkt sich zwar negativ auf das Gewicht des Aufbaus aus, ist aber unerlässlich um eine entsprechende Bewegungsfreiheit zu erlangen. Neben einem regulären Monitor wäre auch ein Multiview fähiges Gerät denkbar. Hier werden mehrere Bilder nebeneinander auf einem Monitor angezeigt. So könnte man eine Stereoskopische Vorschau erreichen, indem die beiden Bilder, wie bei einem Stereoskop, mechanischen voneinander getrennt werden. Zusätzlich ist auf diesem Weg die Kontrolle über beide Kameraeinstellungen gleichzeitig gewährleistet.

Einen anderen Ansatz stellt ein autostereoskopischer Monitor dar. Auf diesem werden die Stereohalbbilder direkt fusioniert und können zur Abschätzung der Tiefenwirkung verwendet werden. Problematisch ist dabei jedoch die Bildverarbeitung. Bisher wurden bei der Recherche nur Monitore mit einer zusätzlichen Verarbeitungseinheit ausfindig gemacht. Diese können über den Monitor eine Live-3D-Ansicht erzeugen. Unter diesen Monitoren befinden sich Geräte, die nur in Analglyphenansicht bzw. mit einer Shutterbrille betrachtet werden können. Ein weiteres Gerät bietet, mit Hilfe einer Video-Kontrolleinheit und entsprechendem Monitor, eine räumliche Wirkung ohne optische Hilfsmittel. Die Schwierigkeiten sind aber erneut die Stromversorgung, das Gewicht und auch die Größe der einzelnen Monitorkomponenten. Auch die unterstützten Eingangssignale sind bei zweien der Geräte, Vidimensio und Dimension Technologies²⁹, nicht für HD-Signale tauglich, müssen folglich vorher konvertiert werden. Computermonitore wie beispielsweise von der Firma SeeReal sind ebenfalls von einem Rechner abhängig, welcher die Bilder der Kameras in vertikale Halbbilder zerlegt. Je nach Computer ist dies auch in Echtzeit möglich. Eine Entwicklung der Firma SpectSoft kommt dem bisher verwendeten Aufnahmeserver der Hochschule sehr nahe. Es handelt sich um eine Aufnahme-Vorschau-Kombination. Das 3D Live System besteht aus einem leistungsstarken Rechner mit zwei HD-SDI Eingängen. Intern wandelt die Software die beiden Ansichten zu einem 3D-Bild um, welches in Echtzeit ausgegeben werden kann und die Kontrolle der Einstellungen zulässt. Auch weitere wichtige Parameter der Stereoskopie können eingestellt werden.³⁰ Der Hersteller arbeitet derzeit an einer portablen Version des Systems. Alternativ wird auch eine Linux compatible Software angeboten, welche allerdings an bestimmte Hardware-Komponenten gebunden ist. Aus Kostengründen würde sich weiter ein System mit SD-TFT-Monitor am besten eignen. Die Einstellungen der Kameras, der Helligkeit und zur Kontrolle des Bildausschnittes können auch ohne hohe Auflösung vorgenommen werden. Einzig für die Überprüfung der Schärfe könnte ein HD-Monitor

²⁹ Siehe Anhang 3

³⁰ Vgl. SpectSoft 08.07.2009

einen Vorteil bedeuten. Die HD-Geräte liegen, je nach Ausstattung, etwa zwischen 500€ und 5000€ Händlerpreis. Geräte im Standard-Definition Bereich können dagegen schon von etwa 140€ bis 1300€ erstanden werden. Die vorhandenen Eingänge des Monitors und die akzeptierten Videoformate müssen jedoch beachtet werden und möglichst auf die Kameras oder Rekorder abgestimmt sein. Ein externer Konverter kann ebenso für ein kompatibles Signal sorgen, muss für den Betrieb allerdings zusätzlich mit Spannung versorgt werden. Auch in diesem Fall könnte ein HD-Monitor einfacher zu handhaben sein.

4.8.2 Sucher

Ein Sucher wie er bei Videokameras verwendet wird, würde der Vorschau eines Monitors entsprechen. Allerdings würde dieser weniger Bewegungsfreiheit für den Kameramann bedeuten. Auch Kamerasucher setzen ein bestimmtes Format und einen speziellen Anschluss voraus, der meist auf eine bestimmte Kamera oder einen Hersteller abgestimmt ist.

Eine Möglichkeit die Vorschau auch in diesem Fall räumlich zu halten, wäre ein stereoskopischer Sucher wie er auf der NAB 2009 vorgestellt wurde. Prinzipiell lässt sich dies einfach durch die Installation zweier Sucher nebeneinander realisieren, wobei dabei das Bild der linken Kamera auf den linken Sucher übertragen wird und umgekehrt. Die auf dem Markt erhältlichen Geräte sind meist nicht mit den relevanten Mini-Kameras kompatibel.

Da ein Kamerasucher auch nur aus einem kleinen LCD besteht, benötigt er genau wie ein Monitor eine Stromversorgung, die im Regelfall direkt von der Kamera übertragen wird. Diese Möglichkeit wird nicht von allen Geräten unterstützt. Für eine stereoskopische Vorschau müssten zwei Kamerasucher eingeplant werden, die knapp nebeneinander montiert werden. Vorteile der Vorschau-Geräte sind die geringere Baugröße und das Gewicht.

4.9 Stromversorgung

Die Stromversorgung bei einem Kameraaufbau, kann bekanntermaßen durch ein Netzteil oder einen Akku realisiert werden. Die Verbindung mit einer Steckdose beinhaltet dabei allerdings immer die Einschränkung der Bewegungsfreiheit und die Wahl der Drehorte, wie auch mit dem bisherigen Produktionssystem festgestellt wurde. Consumer-Kameras besitzen daher in der Regel einen Akku, der die geforderte Mobilität gewährleistet. Auch professionelle Kameras sind mit einer mobilen Stromversorgung ausgestattet. Industriekameras und Minikameras unterscheiden sich allerdings in der Bauweise. Aufgrund der Einsatzgebiete und der Bauformen werden die Geräte meist mit einer externen Stromversorgung gespeist, was für die Konstruktion eines mobilen Aufbaus ein Problem darstellt. Die passende Lösung für eine ausreichende Stromversorgung hängt immer von den Anforderungen des Systems ab. Es muss also

beachtet werden, wie viel Strom, welche Spannung und wie viele Geräte betrieben werden müssen. In einem beispielhaften Fall könnte dies bedeuten, dass zwei Kameras oder Bedienteile, ein Monitor und gegebenenfalls ein Genlock oder Timecode Generator betrieben werden muss. Das Aufnahmegerät bleibt zunächst unbeachtet, da Feldrekorder meist über eine eigene Stromversorgung verfügen. Dies gilt in vielen Fällen auch für den Monitor. Als nächstes sollte die angestrebte Betriebsdauer festgelegt werden, um die Kapazität des Akkusystems zu ermitteln. In Verbindung damit muss auch die Ladezeit der Akkuelemente beachtet werden, da diese sich ebenfalls indirekt auf die Einsetzbarkeit des Systems auswirkt. Die Anzahl der eingesetzten Batterien macht sich auch im Gewicht und der Handlichkeit bemerkbar. Es wäre daher denkbar die Akkus wie einen Akkugürtel anzuordnen, den der Kameraoperator umschnallt. Dies würde das Gewicht des Rigs deutlich reduzieren und die Handhabung vereinfachen. Für eine Eigenkonstruktion muss die Stromversorgung ein- und ausgeschaltet werden können. Ein Überspannschutz sollte integriert werden, damit die beim Einschalten entstehenden Spannungsspitzen den Geräten keinen Schaden zufügen. Eine solche Konstruktion sollte deshalb ein Fachmann vornehmen.

Die unterschiedlichen Akkutechnologien wie Nickel Cadmium, Nickel Metall Hybrid oder Lithium Ion weisen unterschiedliche Merkmale auf, die für die Verwendung mit einem Kameraaufbau nicht unerheblich sind. Nickel-Cadmium-Akkus (NiCd), wie z.B. normale Mignon Batterien, haben die Nachteile, dass sie nur eine geringe Kapazität ermöglichen und bei falscher Ladung, ein Memory-Effekt auftritt.³¹ Dies macht sie ungeeignet für eine Verwendung an der Hochschule.

NiMH (Nickel-Metall-Hybrid) Akkus, wie sie z.B. bei kompletten Akkugurten eingesetzt werden, sind im Vergleich zu NiCd Akkus deutlich Umweltverträglicher und besitzen eine höhere Lebensdauer. Nachteilig ist allerdings die Tatsache, dass sie nur etwa 500 mal wieder aufgeladen werden können und eine starke Temperatureinwirkung besitzen. So sinkt die Leistung bei 0° stark ab und ist bei etwa -20° nicht mehr einsetzbar.³²

Eine weitere, verbreitete Form ist der leichtere Lithium-Ionen-Akku (Li-Ion). Dieser ist weniger von den Außentemperaturen beeinflusst und auch der Memory-Effekt bleibt aus. Die Lebensdauer beträgt ebenfalls ca. 500 Ladungen, allerdings ist auch das Alter der Zelle hierfür von Bedeutung.³³

Gängige, professionelle Kamera-Akkus werden in der Li-Ion-Technologie gefertigt und besitzen meist einen V-Mount oder Anton-Bauer-Goldmount. Akkus, Kameras und Ladegeräte mit V-Mount sind bereits an der Hochschule vorhanden und würden sich deshalb anbieten. Für eine Eigenkonstruktion sind auch Adapterplatten einsetz-

31 Vgl. Nickel-Cadmium-Akkumulator – Wikipedia, 05.07.2009

32 Vgl. Nickel-Metallhydrid-Akkumulator – Wikipedia, 05.08.2009

33 Vgl. Lithium-Ionen-Akkumulator – Wikipedia, 01.08.2009

bar, welche dann eine Verbindung zur Kamera oder dem Bedienteil herstellen und beispielsweise an einen Gürtel oder das Rig montiert werden können. Kapazitäten von etwa 50Wh (Wattstunden) bis ca. 500Wh werden angeboten. Preis und Gewicht nehmen aber ebenfalls mit der Leistung zu. Auch die schon zuvor erwähnten Akkugürtel könnten eine passende Stromversorgung gewährleisten. Denkbare Komponenten wären RED Bricks oder ihr Nachbau von Switronix, welche im Verhältnis zur Leistung einen günstigen Preis erreichen. Weitere Bausteine bieten unter anderem SWIT, Sony, House of Power oder IDX. Die Akkus müssen für den Einsatz jeweils mit Adapterplatte eingebaut werden.

5 Praktische Umsetzung Version 1

5.1 Objektiv

Die Anschaffung von Objektiven, ohne vorher das Kameramodell festgelegt zu haben, ist aufgrund der gewonnenen Erkenntnisse so gut wie nicht möglich. Da für die Konstruktion und Überprüfung des Rigs jedoch trotzdem Objektive besorgt werden mussten, sollte eine möglichst flexible Lösung gefunden werden. Aus diesem Grund fiel die Entscheidung für das Prototyping auf das Kowa LM12JCM Objektiv. Es handelt sich um eine 12mm Festbrennweite. Es wird manuell bedient und ist für einen 2/3" Sensor ausgelegt. Die gewählte Brennweite macht das Objektiv auch in Zukunft verwendbar. Sie wurde so ausgesucht, dass es in Verbindung mit einem 2/3" Sensor leicht weitwinklig wirkt, mit einem 1/2" Sensor der Normalbrennweite entspricht und mit einem 1/3" Sensor im leichten Telebereich agiert. Der Außendurchmesser ist mit 32,5mm sehr gut geeignet um kleine Stereobasen einzustellen. Hinzugerechnet werden muss allerdings noch der Durchmesser des zur Steuerung notwendigen Zahnrads, da dies die einzige zuverlässige synchrone Einstellmöglichkeit bietet.



Abbildung 7 - Kameraobjektiv mit einfachen Zahnrädern

5.2 Rig

Für die Umsetzung des Rigs wurde, aufgrund der Vorgaben, die Side by Side Variante gewählt, da sie klein genug konstruiert werden kann. Die Breite des Rigs beträgt 400mm. Mit diesen Abmessungen sollte problemlos in kleinen Räumen wie Autos oder Aufzügen gefilmt werden können. Die Breite ist außerdem begründet durch die Tatsache, dass es nicht möglich ist zwei Stative näher aneinander zu positionieren. Es wird also eine lückenlose Einstellung der Stereobasis gewährleistet. Die Probleme der Parallelität werden, durch einen auf zwei Führungsschienen laufenden Schlitten, be-

hoben. Auf diese Weise kann keine ungewollte Verdrehung der Kameras stattfinden. Zudem muss die Kamera nach der Veränderung der Stereobasis nicht erneut ausgerichtet werden. Dies spart Zeit und trägt zur Konstanz der Videos bei, was wiederum Zeit und Arbeit in der Postproduktion einspart. Die Stereobasis soll auf Skalen abgelesen werden können, was auch diese Einstellung wesentlich vereinfacht.

Als Material für das Rig wurde Aluminium als Hauptbestandteil und Messing bzw. Silberstahl als Materialien für spezielle Teile vorgesehen. Dies hat den Hintergrund, dass die Konstruktion gegen Korrosion geschützt ist. Außerdem wird das Gewicht möglichst reduziert. Gegenüber Kunststoff bietet Aluminium eine höhere Festigkeit und Langlebigkeit. Die Führungsstangen werden aus Silberstahl gefertigt, da sie keine Abnutzung zeigen dürfen dabei sehr stabil und ebenfalls rostfrei sein müssen. Die Zahnräder und Zahnstangen hingegen sind aus Messing geplant, da das Material den Leichtlauf des Systems gewährleistet und auch nicht rostet.

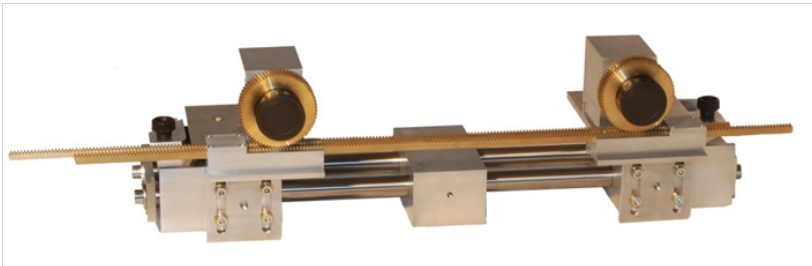


Abbildung 8 - Stereorig

Die Anforderung, dass die Objektive manuell und synchron verstellt werden können, stellte die größte Herausforderung dar. Da die Kameras in ihrer Stereobasis variabel bleiben sollten, die Objektive aber zu jeder Zeit mit einander verbunden sein müssen, gibt es nur wenige Lösungswege. Der Versuch eine Gummibeschichtung zu verwenden, die dann durch Reibung die Objektive verstellt, wurde als nicht praktikabel erachtet, da die Zuverlässigkeit für den Betrieb nicht gewährleistet ist. Als Alternative wurde eine Konstruktion mit Zahnrädern und Zahnstangen entworfen. Die Zahnräder werden auf das Objektiv angepasst und dort befestigt. Die Zahnstange, die sich absenken lässt, verbindet die beiden Objektive und wird von Hand bewegt. Durch das Senken können die Kameras zusammen oder auseinander geschoben werden, ohne dass sich die Einstellung ändert. Beim wieder einrasten muss allerdings darauf geachtet werden, dass die Verzahnung die gleichen Abstände hat. Ist dies nicht der Fall könnte sich die Position einer der beiden Objektivparameter verändern und es wäre keine identische Einstellung mehr vorhanden. In der Praxis könnte der Vorgang so vonstattengehen, dass zuerst die Objektive in die Endposition der Schärfe, Blende und Zooms gebracht werden. Nun wird die Zahnstange abgesenkt und die gewünschte Stereobasis grob eingestellt. Dann bringt man die Zahnstange

vorsichtig wieder zurück in Position und überprüft dabei ob die Zähne an der richtigen Stelle stehen. Ist dies nicht der Fall muss die Stereobasis so geändert werden dass die Einstellung korrekt ist. Da es sich hier nur um wenige Millimeter handelt hat diese Korrektur keinen besonderen Einfluss auf die berechnete Scheinfensterweite. Eine zweite Lösung kann über die Stativplatten realisiert werden. Diese sind nach dem lösen der Klemmschrauben nach Hinten verschiebbar. Die Zahnräder, die ebenfalls verschoben werden, geben die Zahnstange frei und nun kann die gewünschte Stereobasis eingestellt werden.

Das Absenken der Zahnstangen ist auch für die Einstellung der Konvergenz notwendig. Verdreht man die Kameras auf dem Rig sind auch die Zahnräder an den Objektiven verdreht. In diesem Fall können die Zahnräder nicht mehr in die Zahnstange greifen. Bei diesem Vorgang muss also in Kauf genommen werden, dass die Objektiv-einstellung nicht mehr synchron vorgenommen werden kann. Da Konvergenzen allerdings erst für den Nahbereich und Makrobereich verwendet werden, ist auch dieser Umstand nicht von großem Nachteil.

Um die Kameras für Panoramaaufnahmen auf zwei separaten Stativen zu befestigen, wurde je eine Stativwechselplatte in das Rig integriert. Es wurden zwei Manfrotto 501PL Stativplatten angeschafft, welche durch eine Klemmvorrichtung in Position gehalten werden. Bei Bedarf können die Kameras ausgeklinkt und auf die an der Hochschule vorhandenen Manfrotto Stative aufgeschoben werden. Der Vorteil an dieser Lösung ist, dass die Kameras selbst nicht bewegt werden müssen. Wenn die beiden Geräte wieder auf das Rig montiert werden, ist die Einstellung von Winkeln und Abständen nicht von neuem erforderlich. Eine Kontrolle ist aber trotzdem sinnvoll.

Das Rig ist außerdem so konstruiert, dass auch andere Kameras und Objektive ohne allzu großen Aufwand angebracht werden können. Die Absenkung der Zahnstangen sorgt hier für einen gewissen Spielraum, so dass auch Objektive mit größerem Durchmesser (bis zu 65mm inklusive Zahnrädern) montiert werden können. Die Zahnräder und Abstände müssen allerdings neu eingestellt werden.

Die Kontrolle über das System soll durch einen Vorschaumonitor gewährleistet werden, welcher an einer Halterung über den Kameras angebracht wird. Diese Halterung soll gleichermaßen als Griff und Stütze dienen. So können die Kameras in verschiedenen Positionen getragen werden.

Eine genaue Verstellung der Objektivparameter lässt die zusätzlich entwickelte Drehvorrichtung zu. Die Zahnräder greifen wie bei den Objektiven in die Zahnstangen ein und Lenken die Verstellung auf nach Hinten versetzte Einstellräder. Dies lässt die präzise Fokussierung zu und verbessert die Handhabung der Objektive.

5.3 Kameratests

5.3.1 Testaufbau

Vorgesehen ist, die Kamerabilder mit Testtafeln für Farbe, Auflösung und Helligkeit in Verbindung mit einem Waveform-Monitor zu bewerten. So kann festgestellt werden ob ein merklicher unterschied zwischen Machinevision Kameras und speziellen

Filmkameras besteht. Die Unterschiede können die Verwendbarkeit in einem Hochschulaufbau stark beeinflussen. Da die Schule leider über keinen Waveformmonitor mit Firewire-Eingang verfügt, wurde die Entscheidung getroffen, die Bilder anhand eines Software-Waveform-Monitors zu prüfen. Durch diese Entscheidung lassen sich die Ergebnisse der beiden Kameras vergleichen. Die erste Wahl fiel auf das Programm Adobe On Location. Es ist Teil des Adobe Premiere-Pakets und für die Überprüfung von Livebildern während der Dreharbeiten ausgelegt. Weitere Nachforschungen ergaben aber, dass das Programm zum Zeitpunkt des Tests noch nicht für High Definition Formate ausgelegt war und diese deshalb auch nicht verarbeiten kann. Da Adobe Premiere ebenfalls über die Möglichkeit verfügt, einen Waveform-Monitor zu simulieren, konnte auf diese Lösung ausgewichen werden.

Mit dem Vektorskop lassen sich die Sättigung und die Phase der Farben darstellen. Um eine gute Farbwiedergabe zu erreichen, muss das übertragene Signal eines Testbildes in den vorgesehenen Grenzen liegen. Diese Grenzen werden auf dem Vektorskop durch kleine Kästchen dargestellt. Genau in diesen Markierungen muss der entsprechende Farbwert liegen um ein korrektes Signal wieder zu spiegeln. Dieser Idealfall wird allerdings nur direkt von einem Farbbalkengenerator erzielt und ist in der Praxis kaum zu erreichen. Ist es der Fall, dass die Punkte an einer anderen Stelle als den vorgesehenen Kästchen liegen, ist die Farbe entweder nicht genug oder zu sehr gesättigt. Eine weitere Möglichkeit ist eine Phasenverschiebung des Signals, der wiederum eine Farbverfälschung zu Grunde liegt. Auch der korrekte Weißabgleich kann mit dem Vektorskop überprüft bzw. eingestellt werden. Hierbei sollte das Signal so auf dem Monitor abgebildet werden, dass alle Linien im Mittelpunkt des Kreises liegen und einen Punkt formen.³⁴

Mit der Waveformansicht kann die Amplitude des Signals gemessen werden. Diese entspricht der Helligkeit einer Komponente des Signals. Die Darstellung kann Luminanz allein oder Luminanz und Chrominanz beinhalten. Der Wertebereich liegt bei einem Volt von Spitze zu Spitze (Peak to Peak). Deshalb werden die Werte dazwischen in Milli-Volt (mV) gemessen.³⁵

Um die Ergebnisse vergleichbar zu halten, wurde die Beleuchtung der Mess-Tafeln am Messplatz auf 1200 Lux eingestellt, wie zum Beispiel auch bei den Tests des Internet Portal Slashcam. Die momentan im Studio installierten Lampen liefern in etwa 2200 Lux. Gemessen wird die Helligkeit mit dem im Studio vorhandenen Minolta Chromameter. Um die geplanten 1200 Lux zu erreichen wird nur eine der beiden installierten Leuchten verwendet und der Abstand entsprechend angepasst. Die Farbtemperatur liegt bei 4900 Kelvin. Um die Vergleichbarkeit der Tests zu gewährleisten, sollte Helligkeit und Farbtemperatur bei jedem Test identisch eingestellt, zumindest aber überprüft werden.

34 Vgl. Tektronix

35 Vgl. Tektronix

Für die Low-Light Messung soll eine Helligkeit von 12 Lux gewährleistet sein, was durch abgedunkeltes Oberlicht erreicht werden kann. Die Farbtemperatur unterscheidet sich allerdings etwas, da sie in diesem Fall bei 3620 Kelvin liegt. Hier muss ein neuer Weißabgleich vorgenommen werden. Da die Bewertung der Lowlight-Tests nur das subjektiv wahrgenommene Bildrauschen beinhalten soll ist der Unterschied nicht von schwerwiegender Bedeutung.

Als Testbilder werden die vorhandenen Farbcharts der Hochschule und ein Auflösungschart nach ISO-12233 Norm verwendet. Die Farbcharts entsprechen nicht dem 16:9 Seitenverhältnis, welches von der Iconix Kamera geliefert wird, stellen aber die einzige Möglichkeit zur Überprüfung dar. Desweiteren wird ein Testbild in normalen und dunklen Kunstlichtverhältnissen und diverse Außenaufnahmen unter natürlichen Lichtverhältnissen aufgezeichnet. Das Testbild setzt sich aus mehreren Elementen zusammen. Enthalten sind Teile der Farb- und Auflösungscharts, eine Pflanze und eine Puppe. Mit diesen Elementen kann die natürliche Wirkung des Videomaterials subjektiv bewertet werden. Die Außenaufnahmen beinhalten horizontale und vertikale Kameraschwenks, um eventuelle Bildartefakte und Bewegungsunschärfe festzustellen.

Das Aufzeichnungssystem besteht für die Iconix HD-RH1 aus dem bisher verwendeten Aufzeichnungsserver mit den eingebauten Blackmagic Decklink HD-Extreme Videocapture-Karten. Das eingebaute Raid-System, aus drei Festplatten erreicht eine Schreibgeschwindigkeit von etwa 200MB pro Sekunde. Für die AVT Stingray 125C wird, anstelle der Decklink Karten, eine Firewire- Karte verwendet. Diese zeichnet sich durch zwei getrennte Busse aus, die beide über eine Bandbreite von 800MBit pro Sekunde verfügen. Hierdurch wird gewährleistet, dass beide Kameras zur selben Zeit, ohne Bandbreitenverlust, betrieben werden können. Dies ist für den stereoskopischen Betrieb unerlässlich.

5.3.2 Technische Tests

5.3.2.1 Formate

Für die AVT Stingray 125C können Videoformate in RAW, unkomprimiert oder mit Chroma-Subsampling gewählt werden. Die Anforderungen an das Video limitieren die Auswahl allerdings. Die Bandbreite der Firewire-Schnittstelle lässt unkomprimierte Videodatenströme nur in geringer Auflösung oder niedriger Bildwiederholrate zu. Kodierung soll für das System vermieden werden und kann deshalb von vorn herein ausgeschlossen werden. Das Subsampling ist die praktikabelste Lösung, auch wenn die höchste Auflösung der Kamera (1292x964) nicht die geforderte Framerate von mindestens 25fps liefert. Die nächste geringere Auflösung von 1024x768 kann jedoch ein nutzbares Videosignal bieten. Eine höhere Bandbreite ist für diese Kamera nicht zu erzielen, da am Gerät nur ein einziger IEEE1394b Anschluss vorhanden ist. Die Datenrate liegt für 1292x964, 10bit, 25fps bei 622,75Mbit/s, was die effektiv nutzbare Bandbreite von 520Mbit/s deutlich überschreitet. Auffallend ist, dass die

Kamera im Gegensatz zur Iconix ein 4:3 Seitenverhältnis aufweist.

Bei der Iconix HD-RH1, ist eine große Auswahl an Formaten verfügbar.³⁶ Hier wird das Spektrum durch das momentan verwendete Equipment beschränkt. Für die Aufzeichnung von unkomprimiertem Video (4:4:4) wäre eine Dual-Link Verbindung erforderlich, die mit den installierten Decklink-Capture-Karten nicht realisiert werden kann. Die Karten sind anschlussbedingt nur für eine Single-Link-Verbindung ausgelegt. Das hat auch zur Folge, dass keine 1080p Formate unterstützt werden. Das Potential der Kamera kann deshalb nicht voll ausgeschöpft werden. Hinzu kommt, dass die Aufzeichnungsgeschwindigkeit des eingebauten Festplattenraids nicht ausreichen würde diese Datenmengen zu verarbeiten. Da die Bandbreite der HD-SDI Verbindung groß genug ist, können folgende Formate mit dem bestehenden System aufgezeichnet werden:

576i50 (16:9), 720p50, 720p59.94, 1080i50, 1080i59.94, 1080PsF25, 1080PsF23.98

5.3.2.2 Auflösung

Um die Auflösung zu messen, wird ein ISO-12233 Testchart verwendet. Hier muss beachtet werden, dass bei der Auflösung zwangsläufig auch das Objektiv der Kamera einen Einfluss auf die Qualität des Ergebnisses hat. Die HD-RH1 benötigt Objektive die für 3CCD-Kameras hergestellt sind um optimale Ergebnisse zu erzielen, wohingegen die Stingray als 1CCD-Kamera konstruiert ist und sich deshalb die Beschaffenheit des Objektivs von der oben genannten unterscheidet. Durch diese Unterschiede ist es nicht sinnvoll die Objektive von einer Kamera zur anderen zu wechseln.

Für den Vergleich der Kameras muss sich das Testbild genau in den dafür vorgesehenen Bildausschnitts-Markierungen befinden. Da es sich bei beiden mitgelieferten Objektiven um Festbrennweiten handelt, kann nur der Abstand der Geräte zum Chart angepasst werden, um den Bildausschnitt zu verändern. Ebenso muss die Scharfstellung des Bildes genau eingestellt werden. Beide Objektive sind manuell zu bedienen und dementsprechend ist auch die Schärfe zu ziehen. Zur Kontrolle des Bildes steht im Fall der Stingray der Computermonitor und im Fall der HD-RH1 ein Vorschaumonitor zur Verfügung.

Bei der AVT-Kamera fällt auf, dass mit dem 16mm Objektiv ein Wert von nur ungefähr 4 Zeilen (Zeilen/Bildhöhe x 100) noch zwischen den Linien auflöst. Ebenso ist ein starker Moiré-Effekt und chromatische Aberration bemerkbar. Bei der Stingray125C ist es außerdem schwierig die Schärfe des Bildes einzustellen, da man nur auf den PC-Monitor zurückgreifen kann. Die Zoom-Funktion im Vorschau-Player kann hier nur ein klein wenig Hilfe leisten.

Für die Iconix Kamera wird das Testbild bis zu einem Wert von etwa 5 Zeilen noch

³⁶ Vgl. Iconix Video inc. 2007, S. 12

einigermaßen scharf dargestellt. Auch hier wird mit dem 4mm Objektiv ein ausgeprägter Moiré-Effekt bemerkbar, der die Linien überlagert. Außerdem sieht man deutliche chromatische Aberration im Bereich der feinen Strukturen. Die HD-RH1 verfügt über eine Kantenfilter-Funktion, mit der die Schärfe eingestellt werden kann. Diese Einrichtung erleichtert das Scharfstellen deutlich.

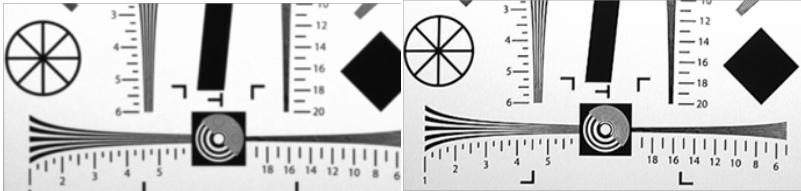


Abbildung 9 - Schärfeschart (li. Stingray, re. HD-RH1)

5.3.2.3 Luminanz

Die Grafik der Stingray zeigt deutlich die Begrenzungs-Funktion der Firewire-Karte oder der Kamera selbst. Der Wertebereich liegt genau bei 0,3V bis 1V. Somit ist gewährleistet, dass die Angaben in Ordnung sind. Ebenfalls kann man erkennen, dass die untere Helligkeitstreppe im Testbild nicht ganz so hell beleuchtet war wie die Obere. Die Beleuchtung war jedoch für beide Kameras und alle TestCharts identisch und verursacht deshalb keine Probleme. Der mittlere Weiß-Bereich liegt bei einem Wert von knapp 1V. Die Schwarzbereiche liegen bei etwa 0,3V. Dies entspricht den optimalen Werten. Die Ergebnisse beider Kameras sind im Allgemeinen nicht sehr weit voneinander entfernt. Im schwarzgrauen Bereich ist der Helligkeits-Betrag der Stingray etwas niedriger, wohingegen die Werte im weißgrauen Bereich etwas höher sind als bei der Iconix.

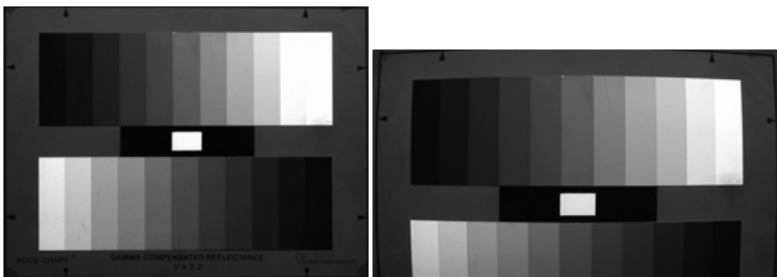


Abbildung 10 - Helligkeitschart (li. Stingray, re. HD-RH1)

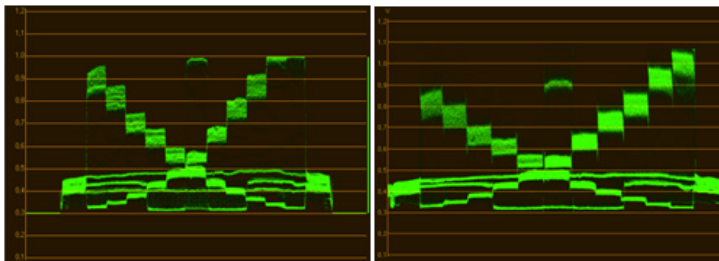


Abbildung 11 – Waveform Helligkeit (li. Stingray, re. HD-RH1)

Im Vergleich muss vor allem die Blendeneinstellung beachtet werden, die für die Stingray bei F6 liegt wohingegen die Blende der Iconix Optik bei F4 liegt. Um das Bild nicht überstrahlen zu lassen, musste das lichtstärkere Pentax-Objektiv an der AVT-Kamera nachgeregelt werden.

Durch das 16:9 Format konnte mit der Iconix nur der obere Bildbereich aufgezeichnet werden. Es ist auffällig dass der Weißpunkt der rechten Seite deutlich über 1V liegt, was theoretisch nicht zulässig ist.

5.3.2.4 Chrominanz

Für die Waveform des gemessenen Bildes der Stingray kann abgelesen werden, dass der hellste Farbbalken, also Gelb, bei knapp über 0,9V liegt. Grau liegt etwas über 0,7V und Hellblau bei 0,75V. Die Weiteren Farben liegen bei 0,6V bzw. 0,55V und 0,45V. Überlagert wird noch der Helligkeitsbereich aus dem Unteren Bildrand, der aber für den Vergleich ignoriert wird, da er bei der Iconix formatbedingt nicht abgebildet wird. Die Werte unterscheiden sich bei der Iconix-Kamera nur um wenige Millivolt.

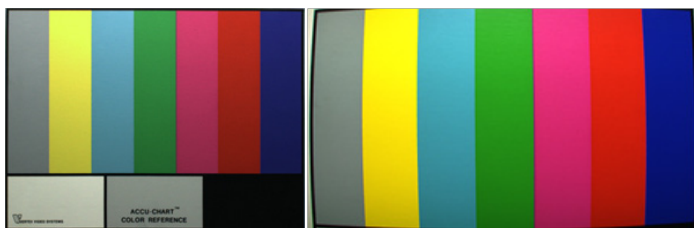


Abbildung 12 - Farbchart (li. Stingray, re. HD-RH1)

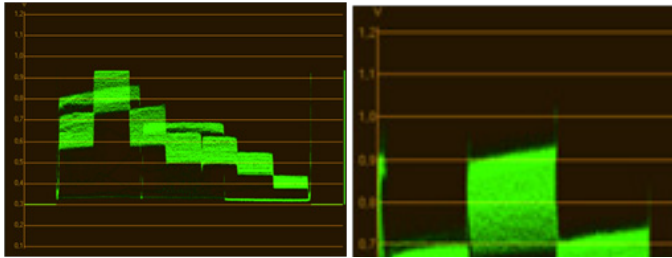


Abbildung 13 - Waveform Farbe (li. Stingray, re. HD-RH1)

Vergleicht man die YCbCr-Diagramme miteinander fällt auf, dass die Helligkeitswerte beider Kameras ebenfalls in etwa gleich ausfallen. Die Farbwerte der Stingray sind in der Cb-Wellenform niedriger, bei ca. 50. Bei der HD-RH1 liegt die Cb-Wellenform zwischen 50 und 60. In der Cr-Wellenform liegen die Werte sehr dicht beieinander. Dies wird auch durch das RGB-Diagramm bestätigt. Für die AVT-Kamera lässt sich ablesen, dass der Grün-Wert am höchsten ist, wohingegen Blau nur einen Wert von 85 erreicht. Die Iconix hingegen weist hohe Rotwerte auf. Der Grünwert ist nur annähernd so hoch wie bei der Stingray aber immer noch bei einem Wert von ca. 95. Blau ist auch bei der HD-RH1 am niedrigsten. Es muss also festgestellt werden, dass die Farbwiedergabe der HD-RH1 deutlich besser ist als die der AVT-Kamera. Die Werte der beiden Diagramme werden in einer Skala von 1 bis 100 gemessen und sind jeweils proportional zu den jeweilig enthaltenen Komponenten. Eine bestimmte Einheit wird laut der Firma Adobe nicht verwendet. Aus diesem Grund können die Ausschläge der Diagramme auch nicht mit denen der Vectorskop-Ansicht verglichen werden.

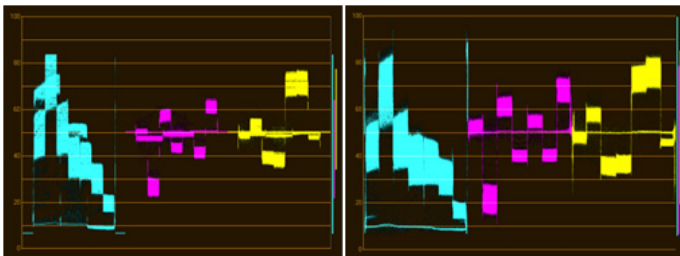


Abbildung 14 - Waveform YCbCr (li. Stingray, re. HD-RH1)

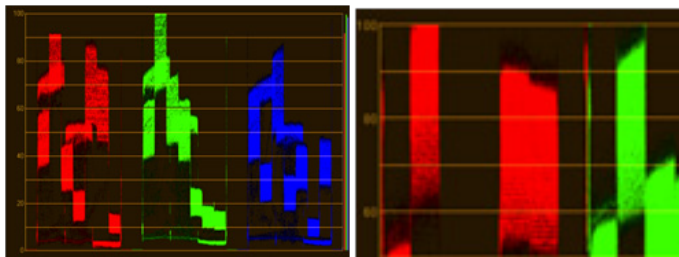


Abbildung 15 - Waveform RGB (li. Stingray, re. HD-RH1)

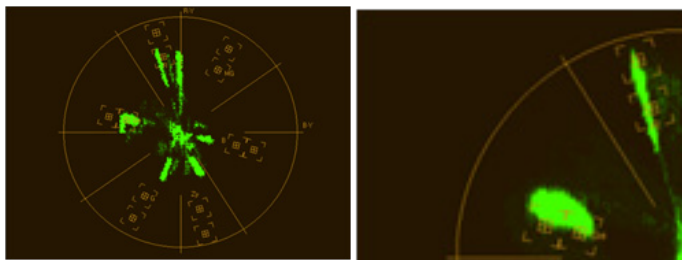


Abbildung 16 - Vektorskop (li. Stingray, re. HD-RH1)

Für die Stingray lässt sich in der Vektorskop-Ansicht einfach erkennen, dass die Phase von Magenta nicht im gewünschten Bereich liegt. Die Sättigungswerte sind auch alle nicht in den regulären Zonen, was den Eindruck der blassen Farben bestätigt. Im Gegensatz zur HD-RH1 kann man erkennen, dass die Farbwerte wenig gestreut sind was wiederum bedeutet, dass mehr reine Anteile der Farbe vorhanden sind. Das kann auf eine bessere Qualität des Objektivs hinweisen. Bei der HD-RH1 lässt sich im Vektorskop ebenfalls eine Phasenverschiebung der Farbwerte erkennen. Magenta ist wie bei der Stingray zu Rot hin verschoben und der Cyan-Anteil ist leicht zu Blau orientiert, wobei hier zusätzlich die starke Streuung der Farben erkennbar ist. Grün erreicht auch nicht vollständig den erwünschten Sättigungsgrad. Insgesamt sind die Werte näher an den Zielfeldern für gesättigte Farben als die Ausschläge der Stingray, was die höhere Farbsättigung nachweist.

5.3.2.5 Handhabung

Der Anschluss beider AVT Kameras an den Aufzeichnungsrechner funktioniert nach Installation der Treiber problemlos. Um allerdings die gesamte Bandbreite der Firewire 800 Schnittstelle nutzen zu können muss, sofern Windows XP mit Servicepack 2 bzw. SP3 installiert ist, der Gerätetreiber der Schnittstelle auf Servicepack 1 zurückgesetzt werden. Das Downgrading ermöglicht es, die auf 400Mbit/s limitierte Bandbreite des SP3 Systems auf 800Mbit/s zu erweitern.

Bei Verwendung der mitgelieferten Software Smartview, ActiveCamViewer oder der Directshow Filter können alle Einstellungen im Menü vorgenommen werden und sind leicht zugänglich. Es besteht auch die Möglichkeit selbst eine Anwendung zu schreiben. Hierzu wird vom Hersteller ein spezielles SDK mitgeliefert.

Als Zubehör zur Kamera wird neben dem 9Pol-Firewire-Kabel nur die Firewire-Karte (IEEE1394b) mit zwei Bussen und eine spezielle Stativplatte benötigt. Alternativ können auch zwei separate Firewire-Karten mit je einem Bus zum Einsatz kommen. Die Stromversorgung findet direkt und ausschließlich über das Firewire-Kabel statt. Um größere Kabellängen zu erreichen, können bei der Fiber-Version der Kamera auch Lichtleiterkabel verwendet werden. Ein Computer ist allerdings unerlässlich. Eine Verbesserung der Mobilität findet demnach nicht statt.

Die Abstände der Kameras können bei der Montage auf ein Minimum von 45mm reduziert werden und stören die Bedienung in keiner Weise. Das mitgelieferte Objektiv beschränkt die Kameraabstände ebenfalls nicht. Die Stereobasis würde nur dann beeinträchtigt werden, falls ein Zoom-Objektiv mit großem Durchmesser verwendet wird und die Abmessungen der Kamera überstiegen werden. Im Allgemeinen wird die Kamera komplett per PC-Software gesteuert und eingestellt.

Bei der Iconix Kamera ist das Menü ebenfalls sehr übersichtlich und alle Einstellungen sind einfach zu finden. Die Steuerung über die Kontrolleinheit ist komfortabel. Einstellungen wie Videoformat und Farbbalken können ohne Menü direkt an den Drehreglern der CCU eingestellt werden. Durch den direkten Anschluss eines Vorschauemonitors sind alle Einstellungen einfach zu überwachen. Es sind Anschlüsse für zwei HD-SDI Verbindungen, sowie Komponentenausgang und Genlock-Eingang vorhanden. Das Objektiv kann bei Bedarf zur Steuerung an die CCU angeschlossen werden.

Zubehör wird nur in Form von Kabeln, Netzteil und der CCU benötigt. Für eine Synchronisation zweier Kameras ist ein externer Genlock-Generator notwendig. Dieser wird aber durch die Aufzeichnung per Direct Show hinfällig und findet erst Verwendung wenn auf einem anderen Gerät gespeichert wird.

Die Montage auf einem Stativ ist problemfrei, da die Kamera selbst über eine Stativbefestigung verfügt. Für einen stereoskopischen Einsatz müsste aber dennoch eine Adapterplatte hergestellt werden, die eine Verdrehung der Kamera verhindert. Die Abstände zweier Kameras sind wie auch bei der Stingray nur durch ein Objektiv limitiert. Die Kamera selbst erfüllt den Mindestabstand von 65mm und könnte sogar eine Stereobasis von 35mm erreichen. Da die komplette Bedienung über die Kontrolleinheit vorgenommen wird ist diese, sofern ein Monitor angeschlossen wurde, nicht durch die Montage behindert.

5.3.2.6 Synchronität

Mit einer Regieklappe und zwei Standard-Softwarerecordern (z.B. Active-Cam-Viewer) lässt sich für die AVT-Kamera sehr einfach eine Synchrone Aufnahme erstellen. Die Clips liegen dann so lange auseinander, wie man Zeit benötigt um beide Aufnahme-Knöpfe zu betätigen (ca. eine halbe Sekunde)

Die Bandbreite der beiden IEEE1394b Verbindungen reicht bei 25fps in der Auflösung 1024x768 aus, um flüssig synchron aufzuzeichnen. After Effects ist in der Lage das 4:3 Material anschließend in einer maßgeschneiderten Komposition zu verarbeiten.

Mit den Direct Show filtern von AVT (AVT ActiveCam Capture Source Filter, AVT_Convert) kann ebenfalls aufgezeichnet werden. Um beide Kameras separat anzusprechen, muss zuerst noch einer der Filter angepasst werden. Die Funktionsweise der Aufnahme ist identisch zu der mit den Blackmagic Karten. Die Einstellungen für Farbe und Helligkeit werden über die Property-Einstellungen des „AVT 1394“-Capture Filters eingestellt. Hier kann auch der Weißabgleich vorgenommen werden. Die Format- und Framerate- Einstellungen, sowie die Videoquelle sollte dagegen über die Property-Einstellungen des Ausgabe-Pins ausgewählt werden. Auf diesem Weg ist ebenfalls eine synchrone Aufnahme realisierbar. Die übrigen mitgelieferten Vorschau-Aufnahme-Programme Fireview und Smartview liefern zwar eine weitere Möglichkeit das Bild einzusehen und aufzuzeichnen, allerdings funktioniert dies nicht synchron.

Für die Kamera von Iconix treffen bei Verwendung der Direct Show-Anwendung die gleichen Eigenschaften zu wie bei der Stingray. Aufgenommen wird über die Decklink Karten und die aktuellen Capture Filter. Würde ein Feldrekorder eingesetzt, müsste die Synchronisation mit externen Geräten sichergestellt werden. Je nach Systemaufbau kann ein Timecode-Generator oder ein Genlock-Generator von Vorteil sein. Diese Konstellation konnte in der Praxis jedoch nicht geprüft werden, da weder eine zweite Iconix Kamera noch ein entsprechender Generator zu diesem Zeitpunkt für Tests bereit standen.

5.3.3 Gestaltungstests

Für die Bildgestaltung bei stereoskopischen Produktionen spielen einige Parameter eine wichtige Rolle. Da bisher nicht alle Einstellungen wie bei normalen Filmaufnahmen vorgenommen werden konnten, soll hiermit getestet werden welche neuen Möglichkeiten sich durch die Verwendung der kleineren, Kameras ergeben.

5.3.3.1 Stereobasis

Ein entscheidendes Kriterium für die Produktion von Stereofilm ist die Weite der Stereobasis. Ist sie zu groß können keine Nahaufnahmen durchgeführt werden. Bisher war diese auf 13,8cm beschränkt. Das hatte zur Folge, dass die Scheinfensterweite bei der kleinsten verwendbaren Brennweite von 30mm etwa 4,1m entsprach. Mit solchen

Entfernungen können keine Nahaufnahmen oder Detailaufnahmen realisiert werden, was sich bisher als gestalterisches Problem herausstellte. Während der Testphasen konnte gemessen werden, dass für die AVT-Kameras ein minimaler Abstand von etwa 4,5cm und für die Iconix-Kamera sogar eine Stereobasis von 3,5cm möglich ist. Dies erweitern die Einstellungsgrößen einschlägig. Durch die kleine Stereobasis könnten nun auch kleinere Brennweiten verwendet werden. Da allerdings für jede Kamera nur jeweils ein Festbrennweiten- Objektiv mitgeliefert wurde, konnte diesbezüglich keine Beweisführung durchgeführt werden. Die mitgelieferten Objektive liefern theoretisch:

AVT Stingray 125C

0,7m Scheinfensterweite (4,5cm SB * 16mm BW)

1,1m Scheinfensterweite (6,5cm SB * 16mm BW)

Iconix HD-HR1

0,14m Scheinfensterweite (3,5cm SB * 4mm BW)

0,26m Scheinfensterweite (6,5cm SB * 4mm BW)

5.3.3.2 Einstellungsgrößen

Die zuvor genannten Werte ermöglichen es nun alle gängigen Einstellungsgrößen für einen Film zu realisieren. Einschränkungen der Tests stellt erneut die feste Brennweite dar. Durch diesen Umstand ist die Scheinfensterweite und somit der minimale Objektabstand nur durch die Stereobasis veränderbar, was nicht als sinnvoll erachtet werden kann. Durch die kleinere Stereobasis sind jedoch auch die nahen Einstellungen bis hin zu Details durchführbar. Es muss jedoch darauf geachtet werden, dass keine zu große Parallaxe entsteht. Für den Fall dass Makroaufnahmen gemacht werden sollen, ist beim Entwurf des Rig die Möglichkeit von Konvergenzstellung der Kameras bedacht worden.

5.3.3.3 Subjektive Bildbewertung

Die Auswertung der Kamerabilder erfolgte auf einem Philips CM0800 PC-Monitor, da keine speziell dafür vorgesehenen Vorschau-Monitore im Studio vorhanden sind.

Testbild

Bei der Stingray (Abb. 19 li.) kann man deutlich erkennen, dass es bei dieser Art Kamera schwieriger ist scharf zu stellen. Die Farben wirken im Gegensatz zur HD-HR1 (Abb. 19 re.) erneut etwas ausgewaschen. Dieses Problem könnte aber in der Postproduktion behoben werden, da es sich um ein gut manipulierbares 4:2:2 Video handelt. Das Grün der Pflanze ist, wie auch bei der Iconix, recht dunkel aber nicht unnatürlich. Das Iconix-Testbild ist dagegen sehr scharf und besitzt einen hohen Helligkeitskontrast.



Abbildung 17 - Testbild (li. Stingray, re. HD-RH1)

Lowlight

Da das Objektiv der Stingray lichtstärker ist als das der HD-RH1, konnte hier eine weiter geöffnete Blende eingestellt werden. Das führt zu dem deutlich sichtbaren Helligkeitsunterschied. Für das AVT Bild ist kaum Rauschen über den Philips Monitor zu erkennen. Die HD-RH1 ist nicht besonders lichtempfindlich, zeigt aber trotzdem auch ein relativ geringes Bildrauschen bei schwacher Beleuchtung. Der Gain-Regler wurde für beide Kameras auf das Minimum eingestellt.

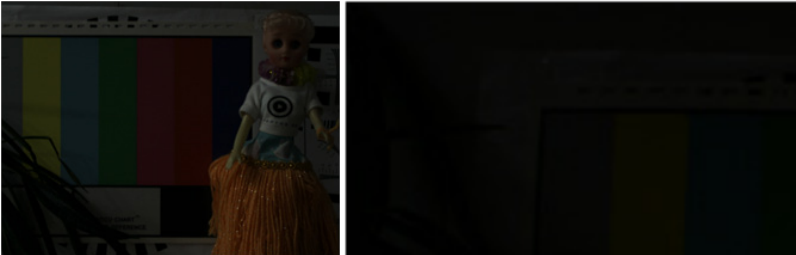


Abbildung 18 - Lowlight (li. Stingray, re. HD-RH1)

Außenaufnahme, horizontaler und vertikaler Schwenk

Stingray 125C

Die Helligkeit des Bildes ist über den Monitor gut einstellbar. Die Farben sind sehr blass und das Bild erscheint fast monochrom. Dies kann einerseits an den schlechten Lichtbedingungen liegen, wird aber auch durch einen nicht optimalen Weißabgleich, der leider erst im Nachhinein festgestellt wurde, hervorgerufen. Auf die Bewertung der Bewegungsartefakte der anderen Bilder hat dies jedoch keinen Einfluss.

Da es schwierig ist, die Geschwindigkeit der Kamerabewegungen exakt gleich durchzuführen, können sie nicht von Kamera zu Kamera verglichen werden. Bei schnellen Bewegungen ist ein Verwischen der Bilder zu bemerken. Es handelt sich aber nur um Bewegungsunschärfe. Artefakte treten keine im Bild auf.



Abbildung 19 - Stingray: Außen, Pan, Tilt

HD-RH1

Für die Aufnahmen entstanden, durch das sehr weitwinklge Objektiv, schwierige Lichtverhältnisse. Es wurden dunkle Vordergrundbereiche und der bewölkte Himmel zusammen erfasst. Trotzdem erreichte die Kamera eine gute Farbwiedergabe und nur die Wolken wirken überstrahlt.

Auch hier ist bei schnellen Bewegungen der Kamera eine starke Bewegungsunschärfe zu erkennen. Artefakte entstehen aber ebenfalls keine.



Abbildung 20 - HD-RH1: Außen, Pan, Tilt

5.3.4 Steuerungstests

Für eine Stereoproduktion müssen beide Kameras identisch eingestellt sein. Um dies zu gewährleisten sind die folgenden Parameter genau abzugleichen.

5.3.4.1 Weißabgleich

Der Weißabgleich kann bei beiden Kameras automatisch oder manuell eingestellt werden. Bei der AVT-Kamera findet dies über die mitgelieferte Software statt, bei der Iconix wird er an der Kontrolleinheit eingestellt. Numerische Werte sind für beide Geräte vorhanden, so dass eine zweite Kamera exakt gleich wie die Erste eingestellt werden kann. Eine Anpassung der Farben in der Postproduktion ist dadurch nicht mehr oder nur in geringem Umfang nötig.

5.3.4.2 Shutter

Auch der Shutter kann Reguliert werden. Hier besitzt die Stingray einen Wertebereich von 1µs bis 62 Sekunden. Die Iconix erreicht Verschlusszeiten von 100µs bis 0,03 Sekunden. Die Werte können ebenfalls genau an den Kameras angepasst werden.

5.3.4.3 Objektivsteuerung

Die Tests zur Objektivsteuerung dienen dazu die Mechanik am Rig besser zu planen. Die ursprüngliche Idee eine Steuerung der Blende und der Schärfe über Reibung zu konstruieren sollte in der Praxis überprüft werden. Hierzu wurden die AVT-Kameras parallel auf der Stereoschiene montiert und die Objektive mit Gummiringen ausgestattet. Dann wurde, wie auch im Entwurf gedacht, ein Stab zur Verstellung der Optiken benutzt. Diese Tests zeigten, dass die Reibung nicht groß genug ist. Durch höheren Anpressdruck konnte zwar eine Verbesserung erzielt werden allerdings ist dieses Konzept nur in gewissem Maß anwendbar. Hinzu kommt dass die Gummierung absolut sauber und fettfrei sein sollte, was im späteren Betrieb nicht möglich ist. Es musste eine andere Lösung gefunden werden, um die Zuverlässigkeit des Systems zu gewährleisten. Die Entscheidung fiel auf den Betrieb mit Zahnrädern. Das Problem der Verzahnung ist jedoch, dass die Kameras nicht ohne weiteres verschoben werden können. Eine Veränderung der Stereobasis kann demnach nur erfolgen, wenn die Zahnräder frei stehen. Das Rig wurde so konstruiert, dass die Kameras nach Hinten verschoben werden können. Die Verbindung wird dadurch aufgehoben und die Kameras sind verschiebbar. An der neuen Position muss der Vorgang rückgängig gemacht werden. Dabei ist vor allem zu beachten, dass die Objektive in einer identischen Einstellung eingerastet werden.

5.3.4.4 Framerate

Für die AVT Kamera sind Bildwiederholraten stufenlos bis zu 30fps möglich. Diese Geschwindigkeit hängt allerdings von der Bildgröße, des Formates und somit der Kabel-Bandbreite und der Festplatten-Geschwindigkeit ab. In den durchgeführten Tests wurde eine Framerate von maximal 25 Bildern pro Sekunde erreicht. Allerdings musste auch hier schon die Bildgröße verkleinert werden. Die volle Auflösung von 1292x964 Pixeln benötigt mehr Bandweite.

Die HD-RH1 dagegen hat fest eingestellte Framerates, die den gängigen Standards entsprechen. Mit der Decklink Capture-Karte konnten allerdings nur 23.98, 24, 25, 50, 59.94 und 60 fps genutzt werden. Dieses Spektrum reicht jedoch aus um die Ansprüche dieses Systems zu erfüllen.

5.3.4.4 Timecode

Beide Kameras liefern keinen Timecode, da dieser immer von einem Rekorder oder einem speziellen TC-Generator erzeugt wird. Dies hat den Nachteil dass der TC nicht ohne Weiteres zur Synchronisation verwendet werden kann. Abhilfe würde ein wie zuvor erwähnter Timecode-Generator schaffen. Mit ihm können z.B. angeschlossene Rekorder und der Startpunkt der Aufnahme synchronisiert werden.

5.3.4.6 Filter

Die Stingray bietet die Möglichkeit, Filter zwischen Objektiv und Kamera einzulegen, was allerdings für ND-Filter keine geeignete Lösung im laufenden Betrieb dar-

stellt. Die zweite Option ist, den Filter wie bei Fotoapparaten üblich, vor das Objektiv zu schrauben, falls ein Gewinde vorhanden ist. Bei der Iconix ist dies ebenfalls die einzige Lösung zur Filtermontage. Bei beiden Kameras ist kein Grau-Filter integriert, was bei hellem Sonnenlicht problematisch werden kann. Eine Nachrüstung wäre daher empfehlenswert.

5.3.4.7 Software

Die Software „Active-Cam-Viewer“ ist sehr leicht zu bedienen. Hier können alle Einstellungen der Kameraparameter vorgenommen werden. Die Werte sind per Schieberegler oder direkt als numerischer Wert einstellbar. Man erhält ein Vorschaubild der Kamera in Echtzeit.

DirectShow funktioniert mit den Firewire-Karten, dank der mitgelieferten Capture-Filter für die Stingray, genauso problemlos wie über die vorhandenen Decklink Blackmagic Videokarten. Über die Filter kann das Format des Videos eingestellt und genau wie bei den bisherigen Direct Show-Anwendungen in eine .avi Datei geschrieben werden. Die Synchronisation funktioniert ebenfalls wie bisher.

Bei der Iconix-Kamera ist für den Stereobetrieb nur die vorhandene Directshow-Anwendung verwendbar. Allerdings wurde während der Tests auch Premiere zum capturen der einzelnen Clips verwendet.

5.3.5 Stromversorgung

Die Stromversorgung der AVT-Kamera findet über das Firewire Kabel statt. Falls ein Desktop-Rechner verwendet wird oder wie bisher der Server, muss dieser allein mit Strom versorgt werden. Abhilfe könnte hier ein Notebook schaffen, allerdings wird von AVT darauf hingewiesen, dass die Stromversorgung eventuell nicht ausreicht und die Kamera dann nicht ordnungsgemäß funktioniert. Bei zwei angeschlossenen Kameras ist die Gefahr einer Unterversorgung doppelt so hoch. Der Strom sollte laut Hersteller mindestens 1,5 Ampere pro Kamera betragen.

Iconix liefert zu ihrer Kamera ein Netzteil, welches die Kontrolleinheit und den Kamerakopf mit Spannung versorgt. Es wäre allerdings auch möglich die Kamera mit einer anderen Stromquelle zu speisen, welche 12V Gleichstrom liefert (Siehe Kapitel 6.1).

5.3.6 Mobilität

Obwohl die Kameras selbst sehr klein und leicht sind, werden sie momentan durch den Aufzeichnungsrechner in ihrer Mobilität eingeschränkt. Die Stingray Fiber hat die Möglichkeit ein normales IEEE1394b Kabel anzuschließen, welches bis zu Längen von 4,5m einsetzbar ist. Eine andere Anschlussart stellt die Lichtleiterverbindung dar, die eine Länge von 100m zulässt. Allerdings sind beide Lösungen nicht optimal für ein mobiles System. Abhilfe könnte nur ein Laptop mit entsprechender Firewire-Karte und Energieversorgung schaffen. Eine andere Möglichkeit ist nicht realisierbar,

da die Kameras ausschließlich über einen Rechner bedienbar sind.

Bei der HD-RH1 wird im bisherigen Betrieb ebenfalls nur eine Kabellänge bis 100m erreicht. Die Kamera und die Kontrolleinheit sind jedoch ein in sich geschlossenes System und könnten auch in Verbindung mit einem anderen Rekorder verwendet werden. Alle Einstellungen können an der Control-Unit vorgenommen werden, sofern ein Monitor angeschlossen ist. Durch den Einsatz eines Feldrekorders könnte hier die Mobilität gewährleistet werden. Welche Geräte geeignet sind wird in Kapitel 6 genauer erläutert.

5.4 Bewertung der Ergebnisse

Die Kameras sind in ihren Eigenschaften größtenteils gleich. Unterschiede machen sich allerdings in der Übertragungsrate, der Bildgröße und der Farbwiedergabe bemerkbar. Vor allem die Übertragungsrate lässt die FireWire-Kamera als nicht sinnvoll für ein mobiles, stereoskopisches HD-System erscheinen. Die Bandbreite reicht nicht aus, um unkomprimiertes HD-Video zu übertragen und kann somit nicht die geforderten Bedingungen erfüllen. Die Bedienung ist bei beiden Kameras einfach zu handhaben. Bei der AVT Stingray wird immer ein Computer zur Einstellung, Vorschau und Aufzeichnung benötigt, was sich ebenfalls negativ auf die Mobilität auswirkt. Dieses Problem existiert bei der Iconix-Kamera nicht. Für die Stromversorgung muss eine individuelle Lösung gefunden werden, da der Netzbetrieb abermals die Beweglichkeit begrenzen würde. Timecode wird von keinem der beiden Geräte generiert. Das führt dazu, dass dieser nicht zur Synchronisation herangezogen werden kann. Mit einem speziellen TC-Generator wäre das zu ändern. Eine zweite Lösung der Synchronisation stellt ein Genlock-Generator dar. Der Nachteil einer externen Gleichschaltung ist, dass diese Geräte ebenfalls mit Strom versorgt werden müssen. Zuletzt fällt beim Vergleich der Kameras auf, dass sich die Baugröße unterscheidet. So könnte mit den Iconix-Kameras eine kleinere Stereobasis eingestellt werden, was vor allem bei Nahaufnahmen einen Vorteil schaffen würde. Der Augenabstand von 6,5cm ist von beiden Geräten problemlos einhaltbar.

6 Praktische Umsetzung Version 2

Das System welches letztlich verwendet wurde baute nun auf den vorgegangen Erkenntnisse auf.

6.1 Kaufentscheidung Kamera: AVT GE1910C

Bei der GE1910 C handelt es sich ebenfalls um eine Industriekamera des Herstellers Allied Vision Technologies. Die Kamera wird über die sogenannte GIGE-Schnittstelle³⁷ angeschlossen, die Datenübertragung erfolgt über ein Gigabit-Netzwerk. Die Kameraparameter können entweder über eine mitgelieferte Software verändert werden, ebenfalls über die GIGE-Schnittstelle oder über die I/O Schnittstelle überschrieben werden. Der 2/3 CMOS-Sensor liefert eine Auflösung von max. 1920x1080 bei maximal 32fps und war somit für eine mögliche Nutzung gerade ausreichend. Als geeignete Ausgabeformate kann RGB 8-Bit, RGB 12-Bit, auch in umgekehrter Farbreihenfolge oder YUV 4-1-1, YUV 4-2-2, YUV 4-4-4, genutzt werden. Die Stromversorgung geschieht über ein separates Netzteil, kann aber auch durch einen Akku ersetzt werden. Die Kamera war aufgrund der Schnittstelle und des Preises von knapp 3100€ pro Kamera noch im Rahmen des Budgets. Außerdem waren die schmale Breite der Kamera ausreichend um eine Stereobasis auf dem Rig von 6.5 cm zu erreichen. Die aktuelle Stereobasis im parallelen Aufbau beträgt 6.8 cm. Es lässt sich durch das Austauschen der Objektive noch min. 0.5 cm an Basis verringern, da der Durchmesser der Objektive größer ist als die Breite der Kamera, welches wiederum die Stereobasis der Kameras eingrenzt.

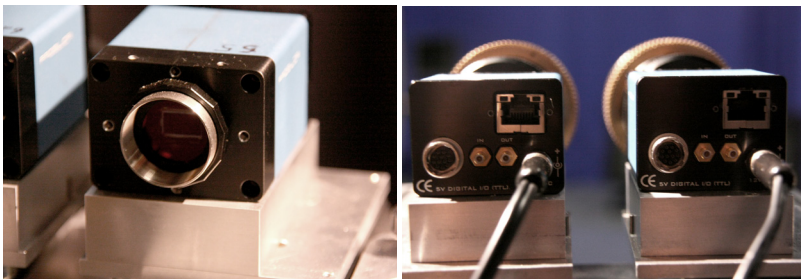


Abbildung 21: Kamera AVT GE 1910C

6.2 Objektiv CVO GMMPZ1664MCN

Die Entscheidung gegen ein Objekt mit Festbrennweite war eine Entscheidung für ein Varifocal-Objektiv mit Zoomfunktion. Da die Größe des Sensors etwas über der

³⁷ http://de.wikipedia.org/wiki/GigE_Vision

Standardgröße für 2/3" Sensoren lag wurde die Auswahl für passende Zoom-Objektive deutlich eingeschränkt, da es bei herkömmlichen 2/3-Objektiven zu Vignetten an der äußeren Rändern gekommen ist. Es konnte jedoch ein geeignetes 1"-Objektiv gefunden werden, welches mit der Brennweite von 16-64mm und einer max. Blendeöffnung von 1.8 entschieden ausreichend war. Das Objektiv hat einen sehr kleinen Schärfebereich und hat gute Markoeigenschaften, welches sehr detailreiche Aufnahmen zulässt. Einen Vergleich mit anderen Objektiven konnte nicht durchgeführt werden, da für die Chipgröße zum aktuellen Zeitpunkt, keine vergleichbaren Zoom-Objektive auf dem Markt verfügbar waren. Preis des Objektives lag bei 800€ und somit ebenfalls im Budget.

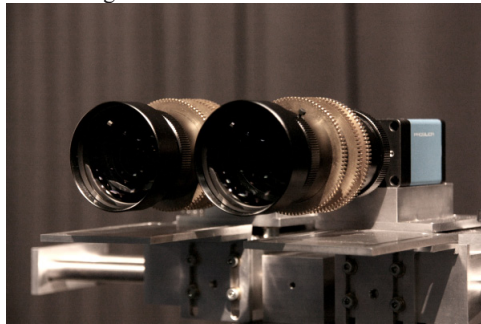


Abbildung 22: Objektiv

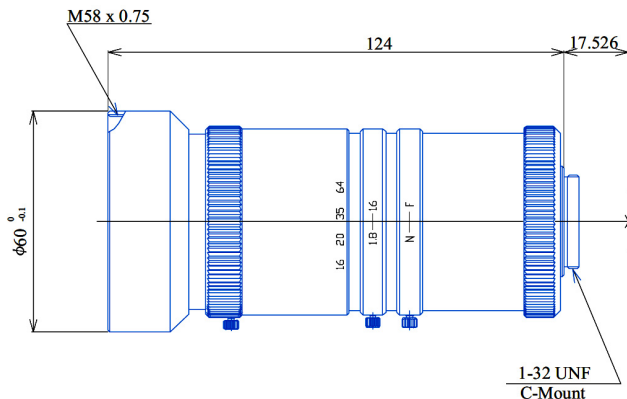


Abbildung 23: Objektiv Schemata

6.3 Aufnahme- Datenübertragung

Zur Übertragung der Bilddaten ist ein Aufnahmerechner mit einer Gigabit-Schnittstelle notwendig. Bei einer stereoskopischen Aufnahme mit 2 Kameras ist es notwen-

dig eine parallele Aufnahme zu gewährleisten. Dazu wurde ein Dual Port Adapter mit 2 Gigabit-Eingängen benötigt. Außerdem sollte der Aufnahmerechner über genügend Leistung verfügen um 2 RAW-Datenstreams übertragen zu können, da es sonst bei hoher Last zu Frameverlust bei der Übertragung kommt.

6.4 Stereobasis

Durch die Kameraauswahl und einiger Optimierungen am Rig war es letztendlich möglich Aufnahmen mit einer Stereobasis von 7cm zu realisieren.

1,12m Scheinfensterweite (7cm SB * 16mm BW)

4,48m Scheinfensterweite (7cm SB * 64mm BW)

6.4.1 Aufnahmerechner-Hardware

Für die Aufnahme der Daten der AVT Kamera wurde eine möglichst mobile Variante eines Aufnahmerechners gewählt. Dabei wurde auf den Erkenntnissen des EcoCast-Projektes aus dem WS 2009/10 aufgebaut, welche einen mobilen Aufnahmerechner entwickelten. Dieser wurde im Nachhinein durch die Erweiterung mit einem Dual Port Adapter für die parallele Aufnahme der beiden GIGE-Streams erweitert.



Abbildung 23: Aufnahmerechner

6.4.1.2 Aufnahmerechner-Hardware - Erweiterungen

Die Aufnahmerechner sollte nur zur Aufnahme beider Streams dienen, eine Kodierung, der Daten sollte nicht bei gleichzeitiger Aufnahme erfolgen, dies sorgt beim Aufnahmerechner zu erhöhter Last auf dem CPU, was wiederum dafür sorgt, dass es zu Framedropping bei der Aufnahme kommt. Im Grunde sollte es vermieden werden, dass parallel weitere Dienste auf dem Rechner laufen, um Framedropping zu vermeiden.

6.4.1.2 Aufnahmerechner-Hardware - GIGE-Schnittstelle

Die GIGE-Schnittstelle ist besonders gut für Datenübertragung über eine lange Entfernung hinweg geeignet. Außerdem können hohe Datenraten bis zu bis zu 1000 MBit/s (basierend auf Gigabit-Ethernet) erreicht werden. Durch die Nutzung der Gigabit-Ethernet-Technologie kann auf herkömmliche Netzkabel zurückgegriffen werden, was auch niedrigere Kosten bedeutet und das System auch sehr ausfallsicher macht. Außerdem waren Kameras mit GIGE-Schnittstelle preislich günstiger als vergleichbare Kamerasysteme mit direktem SDI-Ausgang.

6.4.2 Aufnahmerechner-Software

Für die herkömmliche Aufnahme von einem Stream konnte die mitgelieferte Software genutzt werden. Um jedoch 2 Streams gleichzeitig aufnehmen zu können musste war es notwendig DirectShow zu nutzen. Da der dazu notwendige Input-Filter nicht vorhanden war, musste hierbei auf eine Software von a&b Soft - ActiveGiGe zurückgegriffen werden, die den notwendigen Filter zur Verfügung stellte. Mit diesem Filter ist es nun auch möglich alle notwendigen Kameraparameter direkt zu steuern. DirectShow bietet vor allem noch weitere Möglichkeiten der direkten Bildkontrolle. Es können in Echtzeit Differenzbilder generiert werden, das die Ausrichtung der Kameras zu Beginn deutlich erleichtert. Ebenso lassen sich alle Arten von Codecs direkt mit Einbinden, dies sollte jedoch aufgrund des hohen Rechenaufwands und möglicher Verluste von Frames bei der Aufnahme vermieden werden. Eine Komprimierung kann im Anschluss erfolgen. Weiterhin lassen sich direkt Filter für eine Anaglyphe Echtzeitausgabe erzeugen, welches eine stereoskopische Erstkontrolle auf einem HD-Monitor deutlich vereinfacht. Theoretisch wäre auch eine direkte side-by-side Previewausgabe und eine Ausspiegelung auf einem 3DTV möglich, Filter stellt der Stereomultiplexer zur Verfügung.

6.4.3 Aufnahmerechner-Software - In Betriebnahme

Die erstmalige Inbetriebnahme erfordert einige Systemeinstellungen. Diese sind zum einen das Einstellen von "Jumbo Frames"³⁸ damit die Daten überhaupt vollständig übertragen werden können und es nicht schon beim Stream zum Frameverlust kommt. Außerdem müssen die mitgelieferten Treiber installiert werden und bei Verwendung von 2 Kameras sollte im Vorfeld die IP-Adresse festgelegt werden. Dabei ist allerdings die von a&b Soft mitgelieferte Software "IP-Configurator" hilfreich, da diese die Netzwerkidentifikation übernimmt und man die Kameras nicht von Hand konfigurieren muss.

Die Last des Aufnahmerechners muss immer möglichst klein gehalten werden. Das System reagiert sehr anfällig auf Störungen während des Streamings und es kommt zum Frameverlust während der Aufnahme.

38 http://de.wikipedia.org/wiki/Jumbo_Frames

7 Zusammenfassung

Im Laufe der Arbeit wurde festgestellt, dass es verschiedene Möglichkeiten für alle integrierten Komponenten gibt. Starke Einschränkungen für die Kamera bilden die Datenrate der unkomprimierten Videosignale und die damit verbundenen Übertragungswege. Nur HD-SDI und Gigabit-Ethernet stellen in dieser Beziehung Anschlüsse mit ausreichender Bandbreite zur Verfügung.

Für die Aufnahme des stereoskopischen Videodatenstroms wurde erkannt, dass nur eine vorgefertigte Rekorderlösung in Frage kommt, da alle übrigen Speichermedien zu langsam oder zu umfangreich sind. Es wird meist ein Computer, wie er schon bisher verwendet wurde, benötigt. Das würde die Bewegungsfreiheit nach wie vor enorm einschränken. Eine Akkulösung für die mobile Verwendung stellt jedoch hohe Ansprüche, die nicht einfach zu erfüllen sind. Vor allem das Gewicht und die Unterbringung sind sehr unpraktisch. Je längere Laufzeiten gewünscht sind, desto schwerer wird die Energiequelle.

Selbst unter den speziell gefertigten Feldrekordern genügen nur wenige den Eigenschaften für stereoskopischen Film. Es werden besondere Ansprüche an die Software der Aufnahmegeräte gestellt. Deshalb wurde eine eigene Fieldrecorderlösung mit SSDs und integriertem HD-Monitor entwickelt.

Auf Grundlage der bisherigen und durch diese Arbeit gewonnenen Erkenntnisse, wurde ein Stereorig konstruiert, das die Funktion der synchronen Objektivbedienung ermöglicht und die übrigen Einstellungen für stereoskopische Bildaufzeichnung berücksichtigt. Darüberhinaus konnten noch Konvergenzeinstellung und zukünftige Erweiterungen berücksichtigt werden. Die technischen und gestalterischen Optionen für kommende Projekte wurden dadurch um einiges erweitert. Die Einstellungsparameter des Stereosystems kommen nun einer herkömmlichen Kamera sehr nahe.

Literaturverzeichnis

- Hardware & Drivers - FireWire. Online verfügbar unter <http://developer.apple.com/hardware/drivers/firewire/index.html>, zuletzt geprüft am 21.07.2009.
- Lithium-Ionen-Akkumulator – Wikipedia (2009). Online verfügbar unter <http://de.wikipedia.org/wiki/Lithium-Ionen-Akkumulator>, zuletzt aktualisiert am 01.08.2009, zuletzt geprüft am 10.08.2009.
- Nickel-Cadmium-Akkumulator – Wikipedia (2009). Online verfügbar unter <http://de.wikipedia.org/wiki/NiCd>, zuletzt aktualisiert am 05.07.2009, zuletzt geprüft am 10.08.2009.
- Nickel-Metallhydrid-Akkumulator – Wikipedia (2009). Online verfügbar unter <http://de.wikipedia.org/wiki/Nimh>, zuletzt aktualisiert am 05.08.2009, zuletzt geprüft am 10.08.2009.
- PCI-Express – Wikipedia (2009). Online verfügbar unter <http://de.wikipedia.org/wiki/PCI-Express>, zuletzt aktualisiert am 16.07.2009, zuletzt geprüft am 24.07.2009.
- RAID – Wikipedia. RAID 0: Striping – Beschleunigung ohne Redundanz (2009). Online verfügbar unter <http://de.wikipedia.org/wiki/RAID>, zuletzt aktualisiert am 23.08.2009, zuletzt geprüft am 24.08.2009.
- Andrew Woods (Feb. 1993): Image Distortions in Stereoscopic Video Systems. Stereoscopic Displays and Applications IV. Unter Mitarbeit von Rolf Koch Tom Docherty. Online verfügbar unter <http://3d.curtin.edu.au/spie93pa.html>, zuletzt aktualisiert am 09.03.2009, zuletzt geprüft am 06.08.2009.
- Anja Frank (2007): Stereoskopische HD-Produktion. HS-Furtwangen.
- Eric Schumacher (2008): Analyse der Gestaltungsmöglichkeiten von stereoskopischem unkomprimierten HD-Videofilm in Kombination mit Audio-Wellenfeldsynthese auf Immersionsbasis. Bachelorthesis. Betreut von D. Fetzner N. Hottong. Furtwangen. HS-Furtwangen.
- HDMI Licensing, L. L. C. (Hg.) (2009): HDMI :: Manufacturer :: HDMI 1.4. Online verfügbar unter http://www.hdmi.org/manufacturer/hdmi_1_4/index.aspx, zuletzt aktualisiert am 22.07.2009, zuletzt geprüft am 22.07.2009.
- Iconix Video inc. (2007): HD-RH1 User Manual.doc. Herausgegeben von Iconix Video inc., zuletzt aktualisiert am 27.08.2007, zuletzt geprüft am 18.08.2009.
- Katharina Hörnstein (22. Januar 2005): Ein Kabel für alle Zwecke. Was bei PoE zu beachten ist. Herausgegeben von heise online. (1). Online verfügbar unter <http://www.heise.de/ix/extra/2005/ie0501.pdf>, zuletzt aktualisiert am 06.12.2004, zuletzt geprüft am 21.07.2009.
- Lipton, Lenny (1982): Foundations of the Stereoscopic Cinema, zuletzt aktualisiert am 1982, zuletzt geprüft am 11.05.2009.
- Movie-College.de (2008): Normalbrennweite. Online verfügbar unter <http://www.movie-college.de/filmschule/kamera/normalbrennweite.htm>, zuletzt aktualisiert am 28.07.2008, zuletzt geprüft am 19.03.2009.

- Oliver Röder; Röder, Oliver (2007): Grundlagen der Stereoskopie. Analyse der Aufnahme und Projektion von 3D-Bildern /// Analyse der Aufnahme und Projektion von 3-D Bildern. Saarbrücken: VDM Verlag Dr. Müller.
- PENTAX Europe GmbH: Technik-Tipps - PENTAX Europe GmbH. PENTAX. Online verfügbar unter http://www.pentax.de/de/news/599/foto_technik_tipps.php, zuletzt geprüft am 06.06.2009.
- Ray Zone (2005): 3-D Filmmakers. Conversations with creators of stereoscopic motion pictures. Lanham, Md.: Scarecrow Press (Filmmakers).
- Rick Wilson and John Bercovitz (2003): RMM 3D Encyclopedia - Keystone / Convergence error. Herausgegeben von RMM 3D Encyclopedia. Online verfügbar unter <http://www.rmm3d.com/3d.encyclopedia/keystone/keystone.html>, zuletzt aktualisiert am 21.12.2003, zuletzt geprüft am 10.08.2009.
- SpectSoft (2009): 3D Live. Current "shipping" Feature Set. Herausgegeben von SpectSoft. Online verfügbar unter http://www.spectsoft.com/downloads/3dLive_onesheet.pdf, zuletzt aktualisiert am 08.07.2009, zuletzt geprüft am 16.08.2009.
- Steve Venuti (2009): Introducing HDMI 1.4 Specification Features. Herausgegeben von HDMI Licensing. Online verfügbar unter http://www.hdmi.org/download/press_kit/PressBriefing_HDMI1_4_FINAL_8_0_061809.pdf, zuletzt aktualisiert am 15.06.2009, zuletzt geprüft am 23.07.2009.
- Basic Video Testing -- Vectorscope Techniques > NTSC Video Measurements > Application Note : Tektronix. Tektronix. Online verfügbar unter http://www.tek.com/Masurement/App_Notes/NTSC_Video_Msmt/vectorscope.html, zuletzt geprüft am 18.08.2009.
- The Serial ATA International Organization (2009): Benefits of using SATA. Herausgegeben von The Serial ATA International Organization. Online verfügbar unter http://www.sata-io.org/technology/why_sata.asp, zuletzt aktualisiert am 24.07.2009, zuletzt geprüft am 24.07.2009.
- Trevor Boyer (Feb 18, 2009): Stereo Hype. Stereoscopic 3D is enjoying a renaissance in theaters, but what are its prospects for the rest of the video production market. Herausgegeben von digitalcontentproducer.com. Online verfügbar unter <http://digitalcontentproducer.com/cameras/revfeat/stereo-hype-0209/index.html>, zuletzt geprüft am 25.07.2009.
- Volker Stephan (2008): Konzeption und Implementierung eines praxisorientierten, stereoskopischen High Definition Produktionssystems. Bachelorthesis. Betreut von M. Garcia N. Hottong. Furtwangen. HS-Furtwangen..

Teil 3

Eine Einführung in die Bildverarbeitung von stereoskopischen Views

Eric Schumacher

Abbildungsverzeichnis

Abbildung 1: Bildmatrix	120
Abbildung 2: Bildkonstruktion	123
Abbildung 3: Objektiv	124
Abbildung 4: Prismenblock	125
Abbildung 5: CCD-Verfahren	128
Abbildung 6: FOV und AOV	130
Abbildung 7: Schärfentiefe	132
Abbildung 8: Abbildungsfehler.....	139
Abbildung 9: Kamerarotation	143
Abbildung 10: Konvergenzwinkel	145
Abbildung 11: Zentral- und Parallelprojektion.....	149
Abbildung 12: Lochkameramodell	150
Abbildung 13: Kamerakoordinatensystem	153
Abbildung 14: Epipolargeometrie, konvergent.....	159
Abbildung 15: Achsparalleles System	167
Abbildung 16: Homographie	175
Abbildung 17: Rückwärts-Transformation	189
Abbildung 18: Vorwärtstransformation	189

Tabellenverzeichnis

Tabelle 1: Beispielhafte Videoformate der TV-Produktion.....	121
--	-----

Notation

\mathbf{a} :	Vektor, bzw. Punkt in Vektorform
A :	Punkt, Ausnahme: Bildebene I
\mathbf{a} :	Variable
A :	Matrix
\mathbf{a}^T :	transponierter Vektor (Zeilenvektor)
A^T :	transponierte Matrix (bei $A=(\mathbf{a}_{ij})$ mit $m \times n$ ist $A^T=(\mathbf{a}_{ji})$ mit $n \times m$)
$A^{(-I)}$:	Inverse der Matrix ($A * A^{(-I)} = A^{(-I)} * A = E$) E ist die Einheitsmatrix (quadratisch, sämtliche Werte der Hauptdiagonalen sind Eins, die restlichen Werte sind Null)
$ x $:	Betrag von x
$\ x\ $:	Quadratische Form von x
$\tilde{\mathbf{a}}$:	Vektor in Homogenen Koordinaten
\times :	Kreuz- oder Vektorprodukt zweier Vektoren
$\begin{bmatrix} & \end{bmatrix}_x$:	Zuordnung eines Vektors zu einer schiefsymmetrischen Matrix

1 Einführung

1.1 Das digitale Bild

Im Verlauf dieses Textes werden unterschiedliche Bildverarbeitungsvorgänge, welche entweder kameraintern oder in der Postproduktionsphase durchgeführt werden, beschrieben. Da diese Prozesse mit Hilfe eines Computers durchgeführt werden, ergibt sich als zentrale Voraussetzung für das (Video-) Bild, dass es in einer Form vorliegt, die von einem Computersystem analysiert, verarbeitet und gespeichert werden kann. Aus diesem Grund wird im weiteren Verlauf unter einem Bild ein digitales Videobild verstanden, welches sich auf einem Datenträger befindet, z.B. einer Festplatte, einem optischen Speichermedium oder einem Flash-Speicher. Dies umfasst auch Bilder, die beispielsweise mit einem lichtempfindlichen Sensor erfasst und noch vor der Speicherung auf einem entsprechenden Träger in eine digitale Signalform gebracht wurden. Wird unter dem Begriff "aufgezeichnetes Bild" etwas anderes verstanden als das zuvor Dargelegte, so wird darauf explizit hingewiesen.

Ein digitales Bild besteht aus Elementen, die als Pixel (picture elements) bezeichnet werden und Bildpunkte repräsentieren. Pixel besitzen i.d.R. die geometrische Form von Quadraten oder Rechtecken, wodurch sich ein zweidimensionales Punktfeld ergibt, das in Zeilen und Spalten unterteilt werden kann. Ein Punktfeld mit $M \times N$ Bildpunkten kann demnach durch eine $M \times N$ -Matrix repräsentiert werden. Der Zeilenindex wird als m bezeichnet, der Spaltenindex als n ¹. Der Spaltenindex n erstreckt sich von 0 bis $N-1$, entsprechend läuft der Zeilenindex m von 0 bis $M-1$. Gemäß der Matrixnotation verläuft die vertikale Achse der Zeilen von oben nach unten, die horizontale Achse der Spalten verläuft von links nach rechts². Übertragen auf ein

1 Vgl. Jähne 2005, S. 32.

2 Vgl. Poynton 2003, S. 7.

zweidimensionales Koordinatensystem liegt der Ursprung also in der linken oberen Ecke des Punktfeldes; die Position jedes Pixels kann so durch Matrixnotation angegeben werden.

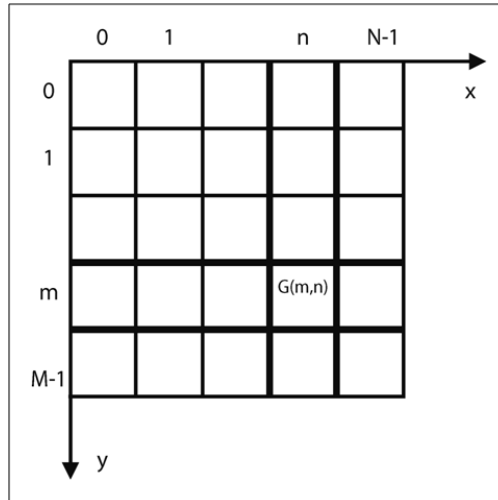


Abbildung 1: Bildmatrix

Aufgrund der Anforderung an die computergestützte Verarbeitung werden in diesem Text hauptsächlich Geräte behandelt, die hochauflösende digitale Videobilder produzieren, indem ein zunächst analog vorliegendes Bildsignal - z.B. aus einem CCD- oder CMOS-Bildsensor - noch vor der Speicherung in ein digitales Signal umgewandelt wird.

Der Begriff digital umfasst im Folgenden die Prozessschritte der korrekten Analog-/ Digitalwandlung eines analogen Videosignals in eine Abfolge von binären Datenwort-Tripel, wobei jedes dieser Tripel die Helligkeits-/Farbwerte eines Pixels an der Position (m, n) repräsentiert ³.

1.2 Digitale Videoformate

Der Begriff Video repräsentiert die Abfolge von einzelnen Bildern mit einer definierten Größe und einem festgelegten zeitlichen Abstand. Ein digitales Videosignal kann vollständig beschrieben werden durch die folgenden Parameter⁴:

³ Vgl. Poynton 1999, S. 5.

⁴ Vgl. Hottong 2009, Folie 136.

- Pixelraster (örtliche Auflösung)
- Zeilenaufbauprinzip
- Bildwechselfrequenz (zeitliche Auflösung)
- Abtastrate
- Quantisierung
- Chromasubsampling
- Kodierungsverfahren

Bezeichnung	SD	HDV 720p	HDCAM SR
Pixelraster	PAL: 720 x 576 NTSC: 720 x 486	1280 x 720	1920 x 1080
PAR	PAL: 1,066 NTSC: 0,9	1	1
Bildseiten- verhältnis	4:3	16:9	16:9
Zeilenaufbau- prinzip	Interlace	Progressive	Progressive
Bildwechsel- frequenz	PAL: 50 Hz NTSC: 60 Hz	24, 25 und 30 Hz	24, 25 und 30 Hz
Abtastrate	13,5 MHz	74,25 MHz	74,25 MHz
Quantisierung	8 bit	8 bit	10 bit
Chromasubsampling	PAL: 4:2:0 NTSC: 4:1:1	4:2:0	4:4:4 oder 4:2:2
Kodierung	Intraframe, DCT/Q/VLC-basiert	Inter- und Intraframe, MPEG 2	Intraframe, MPEG4- StudioProfile
Datentransferrate pro Sekunde	25 Mbit	19.7 Mbit	880 Mbit (4:4:4) 440 Mbit (4:2:2)
Speicherplatz pro Minute	217 MB	142 MB	6,6 GB (4:4:4) 3,3 GB (4:2:2)

Tabelle 1: Beispielhafte Videoformate der TV-Produktion
 Quellen: Ascher; Pincus 2007, S. 770f. & Hottong 2009, Folie 137

2 Kameraparameter

Eine Kamera erfasst einen bestimmten Bereich einer dreidimensionalen Szene und bildet ihn auf einer zweidimensionalen Fläche ab. Ein solches zweidimensionales Bild wird Projektion oder Abbildung genannt. Dieses Kapitel gibt einen Überblick über die verschiedenen Kamerakomponenten, deren Funktionen sowie den jeweiligen Einfluss, welchen eine Komponente auf die entstehende Abbildung ausübt. Wie zu Beginn dargelegt, wird hauptsächlich auf Kameras eingegangen, die ein digitales HD-Video signal aufzeichnen können. Bei den nun folgenden Ausführungen ist zu beachten, dass viele der verschiedenen Parameter unmittelbar miteinander zusammenhängen und sich die Änderung eines einzelnen Werts auf andere Parameter auswirkt, was wiederum Einfluss auf die Abbildung der Szene hat.

2.1 Bildbestimmende Kameraparameter

2.1.1 Brennweite

Wird ein Bild mit einer Kamera aufgenommen, erfolgt eine optische Abbildung der Szene auf der Bildebene dieser Kamera. Die Abbildung wird realisiert, indem die Lichtstrahlen ein optisches System durchlaufen und von diesem auf die Bildebene gelenkt werden. Die Modellierung einer optischen Abbildung erfolgt oft über das Konzept einer einzelnen dünnen Sammellinse, wobei die Bezugsgröße für Berechnungen die Mittelebene dieser Linse darstellt⁵. Die optische Achse steht orthogonal zu der Mittelebene und schneidet diese im *Linsenzentrum* H . Die Lichtstrahlen verlaufen in dem genannten Modell von links nach rechts und treffen auf die Linse, von welcher sie gebrochen werden und u.U. eine Richtungsänderung erfahren. Strahlen, die vor der Brechung parallel zur optischen Achse verlaufen, werden als Parallelstrahlen bezeichnet. Nach deren Brechung werden sie als Brennstrahlen bezeichnet. Die sog. Zentralstrahlen hingegen verlaufen durch das Zentrum der Linse und werden daher nicht gebrochen. Bei einer Sammellinse schneiden sich die Brennstrahlen in einem gemeinsamen Punkt, dem Brennpunkt C . Kommt eine Zerstreuungslinse zum Einsatz, lassen sich die Strahlen so modellieren, dass sie sich vor der Brechung in C schneiden. Die Entfernung zwischen Brennpunkt C und dem Linsenzentrum H wird als Brennweite f bezeichnet und in Millimetern angegeben⁶. Je nach Linsenform wird zwischen einer bildseitigen (Sammellinse) oder einer objektseitigen (Zerstreuungslinse) Brennweite unterschieden. Soll die optische Abbildung eines Szenenobjekts beschreiben werden, so kann dies mit Hilfe der Parallel-, Zentral- und Brennstrahlen sowie C und H erfolgen. Ausgehend von einer Sammellinse und damit einer bildseitigen Brennweite entsteht bei der Abwesenheit von Abbildungsfehlern ein einzelner scharfer Bildpunkt auf der Abbildungsebene. Der Bildpunkt findet sich präzise an der Stelle, an der sich die drei o.g. Strahlenarten, die von ein und demselben einzelnen

⁵ Vgl. Böhringer; Bühler; Schlaich 2008, Bd. 2, S. 194.

⁶ Vgl. Ascher; Pincus 2007, S. 142.

Objektpunkt ausgehen, schneiden. Die Ebene, auf der die scharfe Darstellung erfolgt, wird als Schärfebene bezeichnet und ist in diesem Fall identisch mit der Abbildungsebene. Oft treten jedoch Abbildungsfehler auf, welche die scharfe Darstellung beeinträchtigen (siehe nachfolgender Abschnitt Bildfehler). Im Zusammenhang mit der Modellierung der optischen Abbildung spielen zusätzlich weitere Parameter eine Rolle, welche über die folgenden Gleichungen miteinander in Beziehung stehen^{7, 8}:

- Objektgröße y
- Objektweite a
- Bildgröße y'
- Bildweite a'

$$1/f = 1/a + 1/a'$$

Formel 2

$$y'/y = (a' - f)/f$$

Formel 3

$$a'/a = (a' - f)/f$$

Formel 4

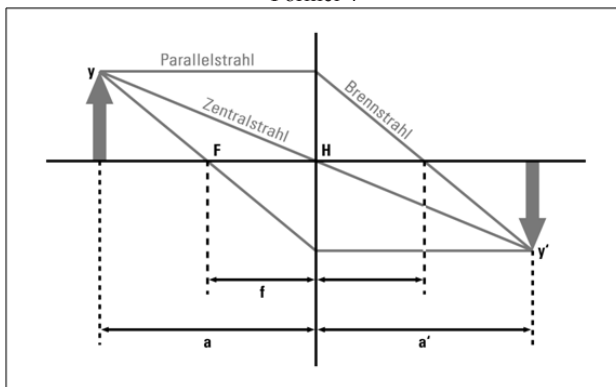


Abbildung 2: Bildkonstruktion

In der Praxis besteht ein optisches System meist aus einer Ansammlung von Linsen unterschiedlicher Formen, die hintereinander angeordnet sind und von denen jede

⁷ Vgl. Meschede 2008, S. 151.

⁸ Vgl. Wikipedia: Linsengleichung.

einen bestimmten Zweck erfüllt (zusammenfassend: Objektiv). Ein Teil der Linsen sorgt für die korrekte Fokussierung der Strahlen auf die Bildebene, d.h. sie ermöglichen eine Anpassung der Schärfenebene. Andere Linsengruppen kompensieren Bildfehler wie z.B. Verzerrungen oder chromatische Aberrationen⁹. Dadurch erweitert sich das Konzept der dünnen Linse, indem sog. Hauptebenen definiert werden, von denen jede eine Gruppe von Linsen repräsentiert¹⁰. Es existieren dann je eine bildseitige und eine objektseitige Hauptebene, das Gleiche gilt für Brennweite und Brennpunkt. Bei Berechnungen für Bildweite und Objektweite stellt die jeweils nächstgelegene Hauptebene und die zugehörige Brennweite den Bezug dar.

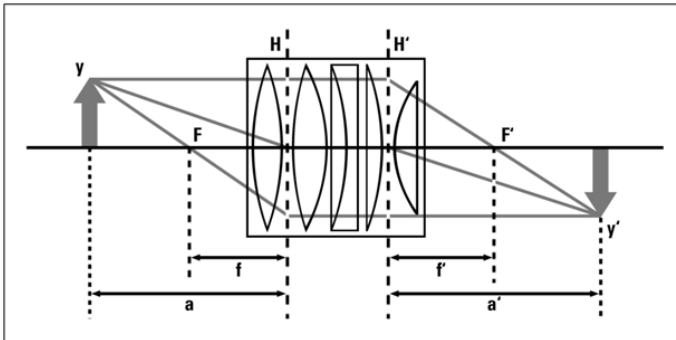


Abbildung 3: Objektiv

Die Brennweite beeinflusst gemeinsam mit dem Faktor der Größe des Bildfeldes Ausschnitt und Perspektive für die zweidimensionale Abbildung der dreidimensionalen Szene. Da die Bildebene i.d.R. eine feste Position besitzt, muss zur Änderung des Bildausschnitts die Position der Linse verändert werden – unter der Voraussetzung, dass weiterhin eine scharfe Abbildung gewährleistet werden kann. Dies wird realisiert durch die Nutzung eines anderen Objektivs oder durch Zoomobjektive, bei denen die Position der verschiedenen Linsen innerhalb des optischen Systems nahtlos verändert werden kann¹¹.

2.1.2 Bildfeld

Dieser Begriff umfasst die Eigenschaften der Komponente(n), auf die die einfallenden Lichtstrahlen auftreffen und das Bild erzeugen. Bei elektronischen Kameras, die auf Magnetbänder, optische Datenträger, Flash-Speicher oder Festplatten aufzeichnen, werden zwischen einem und drei elektronische Bildsensoren eingesetzt, um aus der Energie des einfallenden Lichtes Bildinformationen zu gewinnen¹². Da es hauptsäch-

⁹ Vgl. Schröder; Treiber 2007, S. 80ff.

¹⁰ Vgl. Schmidt 2009, S. 384.

¹¹ Vgl. Böhringer; Bühler; Schlaich 2008, Bd. 2, S. 196.

¹² Vgl. Schreer 2005, S. 39.

lich um die Aufzeichnung von Licht geht, welches im für den Menschen sichtbaren Wellenlängenbereich liegt, werden bei modernen Geräten meist Halbleitersensoren, i.d.R. CCD- oder CMOS-Chips (charged-couple device, complementary metal oxide semiconductor)¹³ eingesetzt.

Bei beiden Sensorarten besteht die Oberfläche eines Chips aus einem feinen Raster von lichtempfindlichen Fotodioden. Genau wie die Bildpunkte werden auch diese Dioden Pixel genannt und können quadratisch oder rechteckig sein. Bei Geräten mit einem Chip werden Farbbilder dadurch erzeugt, dass einzelne Pixel Farbfilter besitzen. Konkret werden durch ein sog. Bayer-Filter mit zwei von vier Pixeln grüne Farbanteile erfasst, mit dem dritten die Rotanteile und mit dem vierten die Blauanteile¹⁴. Die eher selteneren Geräte mit zwei Chips erfassen über ein ähnliches Filtersystem mit dem ersten Chip die roten und blauen Anteile und mit dem zweiten Chip die grünen Anteile. Ein Prismenblock, welcher i.d.R. unmittelbar hinter dem Objektiv platziert ist, lenkt die Lichtanteile dabei auf die entsprechenden Sensorflächen. Modelle mit drei Chips besitzen ebenfalls einen Prismenblock zur Farbseparation, allerdings werden dabei die Rot-, Grün- und Blauanteile getrennt und jeweils einzeln mit der Gesamtfläche eines Sensors erfasst.

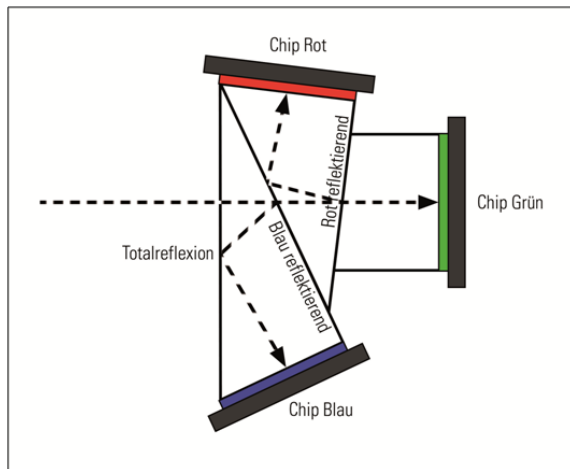


Abbildung 4: Prismenblock

Jede Bauform hat ihre spezifischen Vor- und Nachteile, allerdings bevorzugen einige professionelle Kameralaute Geräte mit einem Chip, da diese keinen Prismenblock besitzen. Der Lichtweg verläuft hier relativ frei von prismatischen Seiteneffekten und

¹³ Vgl. Ascher; Pincus 2007, S. 5.

¹⁴ Vgl. Schröder; Treiber 2007, S. 162.

erlaubt die Nutzung von hochwertigen Wechselobjektiven¹⁵. Ein optisches System, in dem ein Prismenblock zum Einsatz kommt, muss sehr genau an diesen angepasst sein. Die Auswahl an passenden Wechselobjektiven ist daher sehr limitiert¹⁶. Die Größe eines einzelnen Chips und der lichtempfindlichen Pixeldioden sind ein wichtiges Kriterium für die Entstehung des Bildes. Zum Einen werden die örtliche Auflösung und das Seitenverhältnis des Bildes definiert. Zum Anderen wirkt sich die Veränderung der Pixelfläche bei gleichbleibender Sensorfläche auf den erreichbaren Dynamikumfang aus (siehe Abschnitt Bildfehler bei elektronischen Sensoren – Rauschen). Ein Filmstreifen des 35mm Academy Formats ist 35 mm breit, ein Einzelbild weist jedoch Abmessungen von ca. 22 mm × 16 mm auf, bei einem Seitenverhältnis von 1,33¹⁷. Bei Kameras mit Bildsensoren weisen die lichtempfindlichen Sensorbereiche je nach Modell unterschiedliche Abmessungen auf. Deren Bezeichnungen orientieren sich an den Größen von ehemaligen Bildaufnahmeröhren, die früher die Funktion von CCD- und CMOS-Chips erfüllten. Der Durchmesser einer Röhre in Zoll wird zur Angabe der lichtempfindlichen Fläche herangezogen, es ist jedoch zu beachten, dass nicht die gesamte Länge als Bildfenster genutzt werden kann. Ein Chip mit der Bezeichnung 2/3“ weist beispielsweise eine aktive Diagonale von 11 mm auf. Daraus ergeben sich lichtempfindliche Flächen von 8,8 mm × 6,6 mm bei einem Seitenverhältnis von 4:3 und von 9,6 mm × 5,4 mm bei einem Seitenverhältnis von 16:9¹⁸.

Treffen auf einen Sensorpixel Lichtstrahlen auf, übertragen diese durch den sog. inneren Fotoeffekt einen Teil ihrer Energie auf die Sensoroberfläche¹⁹. Die Energie wird zunächst in Form von elektrischer Ladung gespeichert, dabei ist ihre Größe abhängig von Intensität und Dauer des Lichteinfalls. Bei CCD-Chips wird die Ladung von Pixel zu Pixel ähnlich einer Eimerkette weitergeleitet und an einem Ausleseverstärker in Spannung umgewandelt, d.h. es handelt sich um ein serielles Ausgangssignal²⁰. Die Reihenfolge, in der die Pixelreihen ausgelesen werden, bestimmt das Verfahren des Bildaufbaus. Elektronische Kameras unterstützen je nach Modell entweder nur eines der Verfahren oder beide, diese eventuell zusätzlich mit verschiedenen Bildwiederholraten²¹.

Während die Ladung der einzelnen lichtempfindlichen Dioden verschoben wird, darf keine weitere Ladung durch Lichtenergie hinzukommen, um das Bild nicht zu verfälschen (Smear-Effekt). Es gibt unterschiedliche Methoden zum Auslesen von CCD-Sensoren, die verschiedene Ansätze nutzen, um den nachträglichen Lichteinfall zu unterbinden.

15 Vgl. Schmidt 2009, S. 386.

16 Vgl. Ascher; Pincus 2007, S. 137.

17 Vgl. Schmidt 2009, S. 359.

18 Vgl. Schmidt 2009, S. 359.

19 Vgl. Schröder; Treiber 2007, S. 156.

20 Vgl. Meschede 2008, S. 403.

21 Vgl. Ascher; Pincus 2007, S. 100f.

- Full Frame Transfer (kurz: FF): Hier befindet sich eine einzelne CCD-Zeile unterhalb der lichtempfindlichen Fläche. Die Vorgänge der Ladungssammlung und des Auslesens erfolgen getrennt. Bevor das Auslesen beginnt, unterbricht ein mechanischer Verschluss den Lichteinfall für die Dauer des Vorgangs. Die Ladung wird vertikal zeilenweise in jedem Zeilentransfer durch eine horizontale Ladungsverschiebung geleert, wodurch das serielle Signal entsteht²². Neben der Tatsache, dass der mechanische Verschluss relativ aufwändig und stör anfällig ist, stehen im Videobereich für Ladungssammlung und -transport nur ca. 20 ms zur Verfügung. Dies bedeutet, dass die Ladung in den Zellen möglichst lange aufintegriert werden muss, um eine hohe Empfindlichkeit zu erzielen. Daher findet dieses Verfahren verstärkt Verwendung bei Fotoapparaten oder bei Spezialkameras, die beispielsweise in der Wissenschaft oder der Astronomie zum Einsatz kommen.

- Interline Transfer (kurz: IT): Hier liegt der Zwischenspeicherbereich nicht unterhalb, sondern direkt in der Bildwandlerfläche neben den lichtempfindlichen Photodioden. Entstehende Ladungen werden zeitgleich und je Pixel seitlich in die lichtgeschützte Zwischenspeicherzelle geschoben, von wo sie über ein horizontales Transferregister zum Ausgabeverstärker wandern²³. Dadurch, dass der Inhalt jedes Pixels nach der Belichtung in das Transferregister übernommen wird, ist kein mechanischer Verschluss nötig. Die Belichtungszeit kann elektronisch gesteuert werden und darf daher auch sehr kurz sein (elektronischer Shutter, siehe Abschnitt Verschluss). Ein Nachteil dieses Verfahrens ist, dass der Zwischenspeicherbereich einen verhältnismäßig großen Teil der Sensorfläche belegt, wodurch die Dichte der lichtempfindlichen Pixel reduziert wird.

- Frame Transfer (kurz: FT): Ein Chip, der nach diesem Prinzip aufgebaut ist, verfügt über einen lichtempfindlichen Sensorteil sowie einen lichtdicht abgedeckten Speicherbereich nebst einer einzelnen CCD-Zeile als horizontales Ausleseregister. Die einzelnen Ladungen werden nach der Belichtung schnell in den lichtgeschützten Bereich transportiert und während der darauffolgenden Belichtung nacheinander ausgelesen²⁴. Die Abdeckung wird mit einer mechanischen Flügelblende realisiert. Der Ladungstransport muss wesentlich schneller erfolgen als die Belichtung, da sonst der Smear-Effekt entsteht. Für Kameras mit diesem Auslesemodus sind daher kurze Belichtungszeiten nicht empfehlenswert. Beim FT- Prinzip ist die Pixeldichte sehr groß, was zu einer hohen Empfindlichkeit und einem guten Auflösungsvermögen führt. Einen Nachteil stellt die mechanische Blende dar, welche allerdings auch durch einen Verschluss auf der Basis von Flüssigkristallen ersetzt werden kann (siehe Abschnitt Verschluss).

22 Vgl. Schmidt 2009, S. 364.

23 Vgl. Mahler 2005, S. 497.

24 Vgl. Mahler 2005, S. 494.

- **Frame Interline Transfer (kurz: FIT):** Hier werden die Prinzipien von IT und FT kombiniert. Der Speicherbereich ist wie beim IT-Prinzip über Spaltenspalten realisiert, darüber hinaus wird ein FT-ähnlicher Speicherbereich unterhalb des Sensorbereichs genutzt. Die Pixelladungen wandern zunächst schnell in die lichtdichten Spaltenspalten. Von dort aus werden sie in den Speicherbereich transportiert und dann ausgelesen²⁵. Es ist damit keine mechanische Blende nötig, der Smear-Effekt wird stark minimiert und auch kurze Belichtungszeiten sind möglich. Ein Nachteil ist der zusätzliche technische Konstruktionsaufwand.

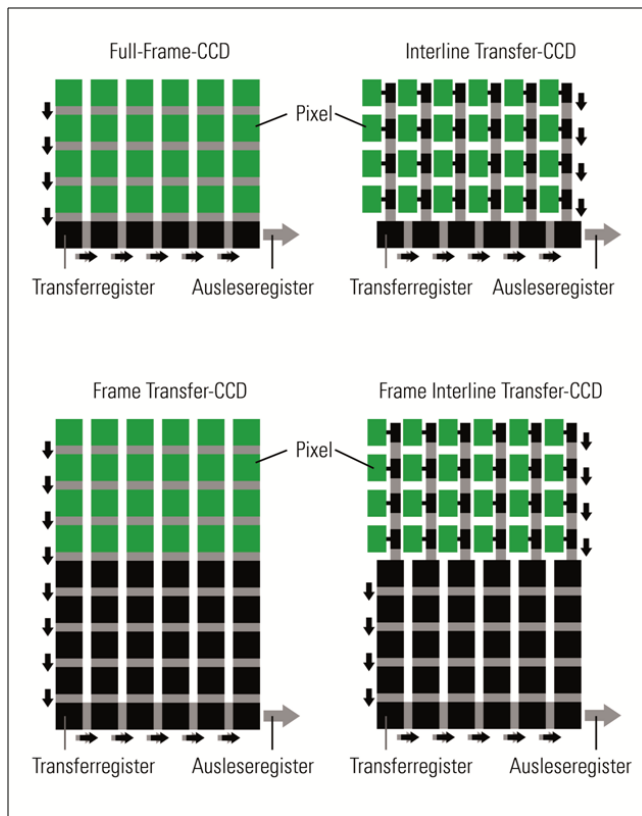


Abbildung 5: CCD-Verfahren

²⁵ Vgl. Schmidt 2009, S. 368.

Bei einem CMOS-Chip befinden sich hinter jedem einzelnen Pixel zusätzliche Transistoren, welche die Ladung direkt in eine messbare Spannung umwandeln²⁶. In diesem Zusammenhang ist auch oft von Active-Pixel-Systemen (kurz: APS) die Rede. Somit können die Pixel ausgelesen werden, ohne dass die Ladungen zuvor verschoben werden müssen. Zusätzlich können einzelne Pixel gezielt getaktet und unabhängig voneinander ausgelesen werden²⁷. Einen kompakten Vergleich liefert Litwiller im Artikel [Litwiller 2005: CCD vs. CMOS: Maturing Technologies, Maturing Markets].

2.1.3 Bildwinkel und Sehfeld

Der Bildwinkel ist ausschlaggebend dafür, wie viele der von der Szene reflektierten Lichtstrahlen auf das Aufnahmemedium gelangen. Bei einem größeren Winkel wird ein größerer Teil der Szene aufgenommen. Werden elektronische Bildwandler mit chemischem Film hinsichtlich der Abbildung einer Szene auf dem Aufnahmemedium verglichen, ist vor allem die Größe der Sensorfläche, bzw. der einzelnen belichtbaren Fläche auf dem Filmstreifen von Bedeutung.

Die Größe dieser Fläche wirkt sich sowohl auf den Bildwinkel (angle of view, kurz: AOV) als auch auf den erreichbaren Schärfentiefebereich aus. Dobbert verwendet zusätzlich den Begriff „field of view“ (kurz: FOV)²⁸. Dieses Sehfeld repräsentiert den sichtbaren Ausschnitt eines Bildes bei einer bestimmten Kombination von Brennweite und Größe des Aufnahmemediums. Zieht man von den vier Ecken der Bildebene je eine Gerade zu den äußersten Bereichen der verwendeten Hauptlinse, entsteht ein Bereich mit der Gestalt eines Pyramidenstumpfes, welcher als Frustum bezeichnet wird²⁹. In der Computergrafik findet sich die Bezeichnung Sichtvolumen³⁰. Die Angabe des FOV bezieht sich auf die Entfernung zwischen den Wandbereichen des Stumpfes und wird in einer Längeneinheit angegeben. Der Bildwinkel wiederum liefert die Information über den sichtbaren Bereich des Bildes mittels einer Gradangabe in Bezug auf die Blickachse.

²⁶ Vgl. Litwiller 2001.

²⁷ Vgl. Schmidt 2009, S. 370f.

²⁸ Vgl. Dobbert 2005, S. 102.

²⁹ Vgl. Dobbert 2005, S. 102.

³⁰ Vgl. Bender; Brill 2006, S. 37f.

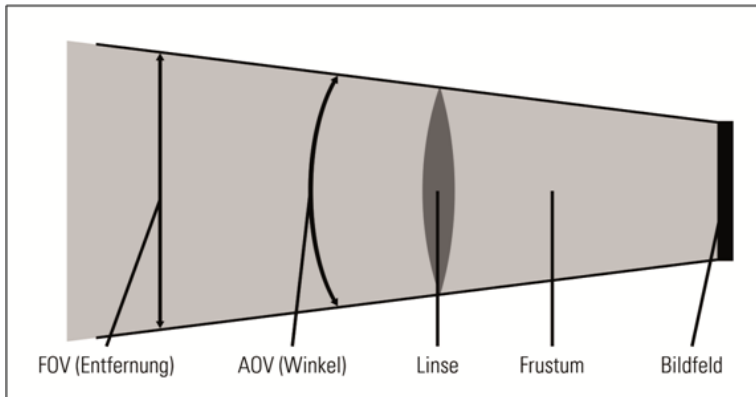


Abbildung 6: FOV und AOV

Da sowohl FOV als auch AOV Informationen über den sichtbaren Bildbereich liefern, werden beide beeinflusst von der Beziehung zwischen Brennweite und Größe des Bildfeldes. Unter der Voraussetzung, dass die Position der Kamera nicht verändert wird, werden FOV und AOV bei gleichbleibender Größe der Bildebene und länger werdenden Brennweiten kleiner, d.h. es wird ein engerer Bereich der Szene aufgenommen; dies würde auch im Falle einer konstanten Brennweite und einer kleiner werdenden Aufnahme­fläche gelten. Umgekehrt wird bei konstanter Größe der Bildebene und einer kurzen Brennweite ein größerer Ausschnitt erfasst, entsprechende Objektive mit einer kurzen Brennweite werden daher oft als Weitwinkelobjektive bezeichnet³¹. Abhängig davon, ob sich der Bildwinkel auf die Breite, Höhe oder die Diagonale des rechteckigen Aufnahmemediums bezieht, handelt es sich um den horizontalen, vertikalen oder diagonalen Bildwinkel. Ohne Spezifizierung ist der diagonale Winkel gemeint, welcher dann mit Hilfe der Brennweite f und der Diagonalen d_B der Bildebene berechnet werden kann³²:

$$\alpha = 2 * \arctan(d_B/2f)$$

Formel 5

2.1.4 Blende

Die Blende wird auch Iris genannt und ist eine mechanische Komponente des Objektivs. Sie besteht aus mehreren Lamellen, die kreisförmig angeordnet sind und sich ineinander verschieben können, wodurch der Öffnungsdurchmesser größer oder kleiner wird³³. Sie erfüllt den Zweck einer sog. Aperturblende, indem sie die Lichtmenge

³¹ Vgl. Schmidt 2009, S. 390.

³² Vgl. Wikipedia: Bildwinkel.

³³ Vgl. Böhlinger; Bühler; Schlaich 2008, Bd. 2, S. 197.

regelt, die durch das Objektiv und auf das Bildaufnahmemedium fällt. Ist die Blende weit geöffnet, kann mehr Licht das optische System durchdringen; bei einer engen Blendenöffnung wird mehr Licht abgeblockt. Wenn die Blende nicht korrekt an die herrschenden Beleuchtungsverhältnisse angepasst wird, sind die aufgezeichneten Bilder entweder über- oder unterbelichtet. Der Status einer Blende wird durch die Blendenzahl k angegeben; sie stellt das Verhältnis dar zwischen der verwendeten Brennweite und dem Durchmesser d_l der Blendenöffnung ³⁴:

$$k = f/d_l$$

Formel 6

Ein Angabe von $k = 2$ besagt beispielsweise, dass der Durchmesser der Blendenöffnung halb so groß ist wie die Brennweite. Die Werte der Blendenzahl sind in Stufen definiert, die sich um den Faktor 2 unterscheiden. Dies liegt darin begründet, dass sich die einfallende Lichtintensität beim Wechsel zum nächsten Stufenwert jeweils halbiert³⁵. Der Öffnungsdurchmesser selbst reduziert sich von Stufe zu Stufe um den Faktor $1/\sqrt{2}$. Die Lichtintensität verringert sich daher mit steigender Blendenzahl. Die Blendenzahl wird oft als Bruchteil der Brennweite angegeben, z.B. $f/5,6$ und dann als fraction stop (kurz: f-Stop) bezeichnet. Der Lichtstärkewert eines Objektivs entspricht dem Kehrwert der kleinsten Blendenzahl (also der größten Blendenöffnung)³⁶. Die Gesamtheit aller möglichen Blendeneinstellungen eines Objektivs wird als Blendenreihe bezeichnet, sortiert von der kleinsten bis zur größten Blendenöffnung ³⁷ (1,0; 1,4; 2,0; 2,8; 4; 5,6; 8; 11; 16; 22; 32). Die Zahlen entsprechen nicht den mathematisch exakten Werten, sondern werden gerundet.

Wird mit hochwertigen Objektiven für analoge oder elektronische Filmkameras gearbeitet, findet sich oft die Alternativbezeichnung transmission stop (kurz: T-Stop). Dieser Wert ist genauer als der f-Stop, denn er bezieht sich nicht auf das angegebene berechnete Verhältnis zwischen Brennweite und Blendenöffnung, sondern auf mit entsprechenden Instrumenten tatsächlich gemessene Lichtwerte ³⁸. Somit werden die Lichtverluste innerhalb des Objektivs selbst, die beispielsweise durch Reflektionen an den einzelnen Linsenkomponenten verursacht werden, berücksichtigt. Würde ein Objektiv existieren, welches keinerlei internen Lichtverlust aufweist, so wären f-Stop und T-Stop und identisch ³⁹.

Neben der Regulierung der einfallenden Lichtmenge kommen der Blende noch weite-

34 Vgl. Ascher; Pincus 2007, S. 149.

35 Vgl. Schmidt 2009, S. 386.

36 Vgl. Cheshire ; Hermstein 1981, S. 46.

37 Vgl. Ascher; Pincus 2007, S. 150

38 Vgl. Schmidt 2009, S. 386

39 Vgl. Ascher; Pincus 2007, S. 151.

re Aufgaben zu, bzw. muss sie bei anderen Faktoren beachtet werden. Die erste Aufgabe besteht darin, Abbildungsfehler im Bild zu reduzieren ⁴⁰, dazu folgt im weiteren Verlauf ein separater Abschnitt (siehe Abschnitt Bildfehler). Der zweite Punkt, bei dem die Blende eine wichtige Rolle spielt, betrifft die Schärfentiefe, die das Objektiv abhängig von der Größe der Bildebene zu erzeugen vermag.

2.1.5 Fokus, Schärfe und Schärfentiefe

Wird mit dem Objektiv auf eine bestimmte Entfernung in der Szene fokussiert, so erscheinen alle Punkte, die sich in dieser Entfernung von der Kamera befinden, scharf. Der Grund dafür ist, dass die Szenenpunkte durch das herrschende Abbildungsverhältnis auch als Punkte auf der Bildebene dargestellt werden. Bereiche, die eine kleinere oder größere Entfernung aufweisen, werden unscharf dargestellt, da ihre Szenenpunkte nicht als Punkte, sondern als Kreise auf der Bildebene dargestellt werden.

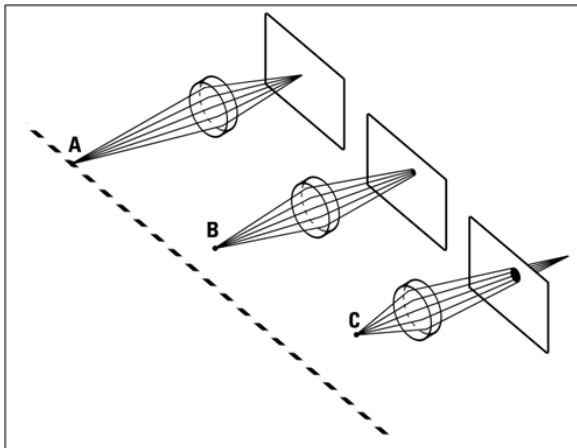


Abbildung 7: Schärfentiefe

Auch die Blende hat Einfluss auf das Abbildungsverhältnis, da sie den Durchmesser des einfallenden Lichtkegels verändert. Bei großer Blendenöffnung ist der Schärfiefenbereich sehr klein. Je weiter die Blende geschlossen wird, desto größer wird der Schärfiefenbereich, der Bereich der scharfen Darstellung vergrößert sich also mit steigender Blendenzahl ⁴¹. Zur traditionellen Bildgestaltung ist oft ein kleiner Schärfiefbereich gewünscht, um einzelne Motive selektiv scharf darzustellen. Die Objekte in anderen Entfernungen bleiben dabei unscharf. Die erreichbare Schärfentiefe hängt zusätzlich von der verwendeten Brennweite ab, welche – wie in den vorherigen Abschnitten erläutert – gemeinsam mit der Größe der Bildebene den Bildwinkel beein-

⁴⁰ Vgl. Dobbert 2005, S. 126.

⁴¹ Vgl. Ascher; Pincus 2007, S. 154.

flusst. Es wurde ebenfalls bereits erläutert, dass die lichtempfindlichen Flächen elektronischer Bildwandler im Gegensatz zu den einzelnen Frames eines Filmstreifens oft wesentlich kleiner sind⁴². Um also mit einer Bildsensor-Kamera den gleichen Bildwinkel, bzw. Bildausschnitt erfassen zu können wie mit einer Kamera, die auf chemischem Film aufnimmt, wird bei identischem Abstand zur Szene eine kürzere Brennweite benötigt. Wenn man jedoch durch eine kleinere Bildaufnahmefläche gezwungen ist, eine kurze Brennweite zu verwenden, wird es wiederum schwerer, gezielt (Tiefen-)Unschärfe einzusetzen⁴³. Viele Hersteller elektronischer Kameras statten daher ihre Geräte für die digitale Cinematografie mittlerweile mit Chips aus, welche in ihren räumlichen Abmessungen an die belichtbare Fläche eines 35mm-Filmbildes angenähert sind, um einen Filmlook mit kleinem Schärfentiefenbereich zu erreichen⁴⁴. Ein Beispiel hierfür ist das Modell RED ONE der RED Digital Cinema Camera Company, die über einen Bildsensor verfügt, welcher bei einer räumlichen Abmessung von 24,4 mm × 13,7 mm eine Anzahl von 4520 × 2540 aktiven Pixel aufweist⁴⁵.

2.1.6 Verschluss

Während die Blende die Lichtmenge regelt, die auf das Bildaufnahmemedium trifft, beeinflusst der Verschluss die Dauer, in der es dem Licht möglich ist, durch das optische System zu fallen⁴⁶. Wenn z.B. ein Filmstreifen belichtet wird, werden im Grunde viele Einzelbilder hintereinander belichtet, i.d.R. mit einer Geschwindigkeit von 24 Bildern pro Sekunde⁴⁷. Zum Belichten wird ein einzelnes Bild hinter das Filmtor transportiert, dort mit einer Druckplatte für die Dauer der Belichtung fixiert und nach der Belichtung weitertransportiert, um für das nächste Bild Platz zu machen. In der Zeit, in welcher der Transport der Einzelbilder stattfindet, soll kein Licht auf das Aufnahmemedium fallen und dafür ist der Verschluss oder Shutter zuständig. Technisch betrachtet kann ein mechanischer Verschluss unterschiedliche Bauformen aufweisen. Bei Fotoapparaten existieren zwei Bauprinzipien, die als Zentralverschluss und Schlitzverschluss bezeichnet werden. Für Filmkameras kommt ein sog. Umlaufverschluss (auch Umlaufblende) zum Einsatz⁴⁸. Dieser besteht i.d.R. aus einer runden lichtundurchlässigen Scheibe, die eine Aussparung aufweist. Die Scheibe rotiert vor der Bildebene und ist mit dem Transportmechanismus des Filmstreifens synchronisiert. Auf diese Weise wird die Öffnung des Filmtores für die Dauer abgedeckt, in welcher der Filmstreifen bewegt wird. Die Verschlusszeit, also die Intervalle, in denen das Filmbild oder der Bildsensor dem Lichteinfall ausgesetzt ist, kann verändert werden, so dass längere oder kürzere Belichtungszeiten möglich sind. Belichtungszeiten werden in Sekundenbruchteilen angegeben, z.B. 1/48 s, 1/50 s oder 1/100 s und sind abhängig von der Bildrate, d.h. der Anzahl der Bilder, die pro Sekunde erzeugt, bzw.

42 Vgl. Schmidt 2009, S. 359.

43 Vgl. Cheshire ; Hermstein 1981, S. 46.

44 Vgl. Ascher; Pincus 2007, S. 136.

45 Vgl. RED Digital Cinema Camera Company (Hg.).

46 Vgl. Doeffinger 2009, 52.

47 Vgl. Dobbert 2005, S. 113.

48 Vgl. Schmidt 2009, S. 425.

aufgenommen werden ⁴⁹. Kameras mit elektronischen Bildsensoren realisieren die Funktion des Verschlusses z.T. über das Ausleseprinzip des Sensors (siehe Abschnitt Bildfeld). Dadurch werden Verschlusszeiten möglich, die mit mechanischen Bauteilen nicht zu realisieren wären (z.B. 1/2000 s). Bei hochwertigen Bewegtbild-Kameras findet sich im Kontext der Belichtungszeit in einigen Fällen auch die Angabe des Verschlusswinkels. Hier kann die Form der rotierenden Scheibe verändert werden, so dass die lichtdurchlässige Öffnung größer oder kleiner wird. Der Wert des Winkels bestimmt dann, wie lange der Verschluss geöffnet bleibt und wann der Verschluss relativ zum Anfang des Frames geöffnet wird, und zwar nach folgendem Verhältnis ⁵⁰:

$$\text{Verschlusswinkel}/(360^\circ) = \text{Belichtungszeit}/\text{Bildrate}$$

Formel 7

Bei einer Bildrate von beispielsweise 25 Bildern pro Sekunde und einem Verschlusswinkel von 180° ergibt sich daraus eine Belichtungszeit von 1/50 s.

Eine besondere Bauform des Verschlusses stellt das von der Firma Arnold & Richter Cine Technik GmbH & Co. Betriebs KG (kurz: ARRI) entwickelte Spiegelreflexsystem dar, bei dem sich auf der Blendscheibe ein Sektor mit spiegelnder Oberfläche befindet. Bei einem 180°-Hellsektor wird so in einer Zeitperiode das Licht je zur Hälfte auf die Bildebene und auf eine Mattscheibe gelenkt, so dass das Bild durch eine Sucherlupe in Echtzeit betrachtet werden kann ⁵¹.

In Bezug auf die Abbildung wird die Helligkeit im Bild also von der Blende und der Verschlusszeit kontrolliert. Darüber hinaus beeinflusst der Verschluss die Menge der Bewegungsunschärfe im entstehenden Bild ⁵². Bei einer längeren Belichtungszeit wird mehr Bewegungsunschärfe sichtbar, d.h. Szenenobjekte erscheinen im finalen Bild abhängig von ihrer Bewegungsgeschwindigkeit mehr oder weniger unscharf. Die Rotationsgeschwindigkeit des Verschlusses zu verändern bedeutet daher, die Belichtungszeit zu verändern. Diese Kontrollfunktion stellt für den Kameramann ein wichtiges Werkzeug zur Bildgestaltung dar.

2.1.7 Optische Filter

Unter diesem Begriff werden Elemente zusammengefasst, die das Bild vor dem Auftreffen auf das Bildaufnahmemedium verändern, jedoch i.d.R. nicht zum eigentlichen Objektiv gehören. Wenn Lichtstrahlen durch ein Filter fallen, so werden sie durch dessen Materialkomponenten und Beschaffenheit beeinflusst. Filter können sowohl vor als auch hinter einem Objektiv liegen, dies hängt meist von ihrer Funktion ab.

- Konversionsfilter (Farbfilter): Mit diesen Filtern können Spektralanteile des

⁴⁹ Vgl. Schmidt 2009, S. 427.

⁵⁰ Vgl. Birn 2007, S. 170.

⁵¹ Vgl. Schmidt 2009, S. 426.

⁵² Vgl. Doeffinger 2009, 30f.

einfallenden Lichts manipuliert werden, indem Lichtstrahlen einer bestimmten Wellenlänge mehr oder weniger stark absorbiert werden ⁵³. Je nach Anforderung kann so ein einzelner Farbtön innerhalb der Szene betont oder abgedämpft werden.

- **Polarisationsfilter:** Im Gegensatz zu Konversionsfiltern absorbieren Polarisationsfilter nicht die Wellenlängen eines bestimmten Spektralbereichs, sondern Lichtschwingungen aus gegenläufigen Schwingungsebenen ⁵⁴. Neben der Nutzung zur Wiedergabe von stereoskopischen Bildpaaren werden Polarisationsfilter auch zu anderen Zwecken eingesetzt. So kann z.B. bei fotografischen Aufnahmen ein Polarisationsfilter genutzt werden, um Kontraste und bestimmte Farbtöne des Motivs zu manipulieren oder zur Entfernung von unerwünschten Reflexionen und Spiegelungen sowohl bei metallischen als auch bei nichtmetallischen glatten Oberflächen (z.B. Fensterscheiben oder Wasser) ⁵⁵.

- **Neutraldichte-Filter (kurz: ND-Filter, Graufilter):** Diese Filter dienen einer gleichmäßigen Reduzierung der Lichtintensität, ohne dabei die Farben zu beeinflussen. ND-Filter weisen je nach Typ unterschiedliche Transparenzen auf, die sich an den Werten der internationalen Blendenreihe orientieren ⁵⁶. Je nach Filterstärke kann so der Lichteinfall entsprechend mehrerer Blendenstufen reduziert werden. Insbesondere wenn die Blende zum Erhalt der Schärfentiefe nicht verändert werden soll, lässt sich der Lichteinfall mit ND-Filtern zu einem gewissen Grad beeinflussen ⁵⁷.

- **Infrarot-Sperrfilter (kurz: IR-Filter):** Ein solches Filter findet sich oft vor dem Bildsensor einer entsprechenden elektronischen Kamera. Die lichtempfindlichen Pixeldioden des Sensors reagieren zu einem gewissen Grad auf das für den Menschen unsichtbare Infrarot-Licht. Da auch viele Lichtquellen einen Teil ihrer Strahlen in Form von Infrarot-Licht aussenden, werden Infrarot-Sperrfilter genutzt, um den Einfall entsprechender Wellenlängen zu verringern, da sich diese störend auf die Abbildungsqualität des Bildsensors auswirken ⁵⁸. Derartige Beeinträchtigungen können z.B. in Form von Unschärfe oder Farbverfälschungen auftreten.

- **Ultraviolett-Sperrfilter (kurz: UV-Filter):** Diese Filterart erfüllt einen ähnlichen Zweck wie ein Infrarot-Sperrfilter, es werden dabei jedoch die Wellenlängen des ultravioletten Spektralbereiches absorbiert.

53 Vgl. Schröder; Treiber 2007, S. 119.

54 Vgl. Böhringer; Bühler; Schlaich 2008, Bd. 2, S. 181.

55 Vgl. Cheshire ; Hermstein 1981, S. 93.

56 Vgl. Schmidt 2009, S. 397.

57 Vgl. Ascher; Pincus 2007, S. 302f.

58 Vgl. Schmidt 2009, S. 397.

- Effektfiler: Neben der grundsätzlichen Verbesserung der Bildqualität existieren Filter, die aufgrund gestalterischer Entscheidungen eingesetzt werden, um bestimmte Bildeffekte zu erzeugen ⁵⁹. Beispielsweise wandeln Sternfilter helle Lichtpunkte im Motiv in Sterne mit vier, sechs oder acht Spitzen um, Spektralfilter hingegen zerlegen Lichtquellen und Reflexionen in mehrstrahlige und mehrfarbige Gebilde. Vignettierungsfiler decken komplette Bildbereiche ab, während bei Diffusionsfiltern feine Details im Bild verschwimmen und Lichtquellen ausgedehnter erscheinen. Darüber hinaus existieren viele weitere Filterarten.

Die Funktion insbesondere vieler Effektfiler lässt sich heutzutage mit digitaler Nachbearbeitung reproduzieren. Für einige Filterarten gilt jedoch, dass sie ihre gewünschte Wirkung nur dann erbringen können, wenn sie bereits während der Aufnahme eingesetzt werden. Dazu zählen z.B. Polarisationsfilter, ND-Filter sowie Infrarot- und Ultraviolett-Sperrfilter.

2.1.8 Bildfehler

Bei der geometrischen Rekonstruktion von Abbildungen wird immer von einem fehlerfreien Abbildungsvorgang ausgegangen. In der Praxis existiert jedoch eine Vielzahl von möglichen Fehlerquellen, welche die korrekte optische Abbildung einer Szene in einem Bild beeinträchtigen können. Im Folgenden werden die wichtigsten Bildfehler erläutert. Es wird zunächst auf Fehler eingegangen, welche innerhalb des optischen Systems durch dessen verschiedene Komponenten und durch das Licht selbst verursacht werden können. Danach werden Fehler erläutert, welche zusätzlich beim Einsatz elektronischer Bildsensoren auftreten können, verursacht durch den Chip selbst oder die ihm nachgeschaltete Signalverarbeitung.

2.1.8.1 Optische Abbildungsfehler

- Öffnungsfehler (sphärische Aberration): Dieser Fehler beruht auf der Brechungseigenschaft und der Krümmung einer Linse ⁶⁰. Abhängig vom Brechungsindex der Linse und dem Winkel, in dem ein Lichtstrahl auf das Linsenmaterial trifft, wird der Strahl in einem bestimmten Verhältnis zu diesem Winkel abgelenkt. Bei einer Änderung des Winkels ändert sich auch die Ablenkung. Einfallende Parallelstrahlen werden in Abhängigkeit von der Entfernung zum Linsenzentrum aufgrund der zunehmenden Krümmung der Linse unterschiedlich stark abgelenkt. Je weiter am Linsenrand der Strahl auftrifft, desto größer wird der Ablenkwinkel. Dadurch kann es passieren, dass sich Strahlen, die von einem gemeinsamen Punkt ausgehen, nicht in einem gemeinsamen Punkt schneiden ⁶¹. Als Resultat wird ein Szenenpunkt unscharf als Kern mit

⁵⁹ Vgl. Ascher; Pincus 2007, S. 302ff.

⁶⁰ Vgl. Meschede 2008, S. 171.

⁶¹ Vgl. Schröder; Treiber 2007, S. 83.

sog. Schleierzonen abgebildet.

- **Koma:** Wie der Öffnungsfehler wird auch dieser Fehler verursacht durch die unterschiedliche Brechung von Lichtstrahlen in Abhängigkeit ihres Einfallswinkels. Hier sind jedoch Strahlen betroffen, die schräg auf die Linse fallen⁶². So werden beispielsweise bei einem Strahlenbündel, das von „unten“ auf die Linse trifft, die unteren Strahlen anders gebrochen als die oberen Strahlen⁶³ (siehe Abb. 8). Der Bildpunkt wirkt dadurch in eine Richtung auseinander gezogen und erscheint als Punkt mit leuchtendem Kern und dreieckig geformten Schleierzonen.

- **Astigmatismus:** Eine Linse kann in ihrer senkrechten Ebene eine andere Brechung besitzen als in ihrer horizontalen Ebene (siehe Abb. 8). So gelten für unterschiedlich einfallende Lichtstrahlen im Grunde unterschiedliche Brennweiten. Auf diese Weise ergeben sich für Strahlen mit unterschiedlichen Eintrittswinkeln verschiedene Abbildungsverhältnisse⁶⁴. Ein Objektpunkt der Szene wird dadurch auf zwei Bildpunkte abgebildet, die sich auf zwei gedachten, hintereinanderliegenden Ebenen befinden, welche parallel zur eigentlichen Bildebene verlaufen. Diese befindet sich dabei zwischen den beiden gedachten Ebenen und der Objektpunkt wird als zwei zueinander senkrechte verlaufende Bildstreifen abgebildet.

- **Bildfeldwölbung:** Dieser Fehler entsteht dadurch, dass die Linse eine abgerundete Gestalt besitzt, die Abbildung jedoch nicht auf einer abgerundeten Fläche entsteht, sondern auf einer ebenen⁶⁵. Daher lässt sich das Bild nicht überall scharf auffangen. Ist der Bereich in der Bildmitte scharf, so nimmt die Unschärfe in den Randbereichen zu und umgekehrt.

- **Verzeichnung:** Bei den bisher genannten Fehlern handelt es sich um Schärfefehler. Bei der Verzeichnung wird die Form des Bildes selbst verändert, unabhängig davon, ob die Abbildung scharf oder unscharf ist⁶⁶. Im Zusammenhang der Verzeichnung ist der Abbildungsmaßstab von Bedeutung. Es handelt sich dabei um das Verhältnis zwischen der Größe des Szenenobjekts und der Größe seiner Abbildung auf der Bildebene. Bei einer maßstabsgetreuen Abbildung werden alle Bildelemente geometrisch korrekt verkleinert, egal an welcher Position sie sich im sichtbaren Bildausschnitt befinden. Viele Objektive verfügen allerdings nicht über eine maßstabsgetreue Verkleinerung, d.h. dass gerade Motivlinien, deren Abbild nicht durch die Bildmitte geht, gekrümmt wiederge-

62 Vgl. Schröder; Treiber 2007, S. 85.

63 Vgl. Meschede 2008, S. 173f.

64 Vgl. Schröder; Treiber 2007, S. 84f.

65 Vgl. Puts 2002, S. 79.

66 Vgl. Ascher; Pincus 2007, S. 169.

geben werden. Der Maßstab verändert sich also mit zunehmender Entfernung von der Bildmitte⁶⁷. Als Resultat dieser Beeinträchtigung treten kissenförmige Verzeichnungen auf, wenn der Maßstab zu den Bildrändern hin zunimmt. Ein abnehmender Maßstab führt zu tonnenförmiger Verzeichnung.

- Farblängsfehler (axiale chromatische Aberration): Die Art und Weise, wie ein Lichtstrahl in einem Medium gebrochen wird, ist nicht nur von seinem Einfallswinkel abhängig, sondern auch von seiner Wellenlänge⁶⁸. So werden z.B. kurzwellige Strahlen (blau) stärker gebrochen als langwellige (rot)⁶⁹. Die Winkelunterschiede bei der Brechung führen zu einer ähnlichen Situation wie beim bereits erläuterten Öffnungsfehler. Strahlen, die von einem gemeinsamen Objektpunkt ausgehen, schneiden sich nicht in einem gemeinsamen Bildpunkt. Stattdessen ergeben sich abhängig von der Wellenlänge mehrere Schnittpunkte und damit mehrere Bilder entlang der optischen Achse⁷⁰. Daraus folgen als Beeinträchtigungen Unschärfe und Kontrastverlust in der Abbildung.

- Farbquerfehler (laterale chromatische Aberration): Die Tatsache, dass Strahlen mit verschiedenen Wellenlängen vom Linsenmaterial unterschiedlich gebrochen werden, führt zu einem zusätzlichen Fehler, welcher die Vergrößerung von Bildpunkten beeinflusst. Unterschiedliche Brechungswinkel entsprechen für die betroffenen Strahlen verschiedenen Brennpunkten, was wiederum unterschiedliche Brennweiten mit sich bringt. Die Brennweite hängt mit dem Vergrößerungsmaßstab zusammen, daher erzeugen die kurzwelligen Strahlen, die von einem Objektpunkt ausgehen, eine kleinere Abbildung als die langwelligen Strahlen⁷¹. Im Bild wirkt sich dies vor allem auf feine Details aus, weiße Bildanteile werden in ihre einzelnen Farben zerlegt dargestellt, schwarze Objekte vor hellem Hintergrund erscheinen mit Farbsäumen versehen.

- Äußere Einflüsse: Fehler im Bild können aus vielen anderen Quellen stammen. Oft handelt es sich um mechanische Beeinträchtigungen in Form von Kratzern, Dellen oder Flecken auf dem Objektiv oder um Staub und andere Fremdkörper, die in den Strahlengang gelangt sind. Weiterhin können sich Objektiveigenschaften durch extreme äußere Bedingungen wie extreme Umgebungstemperaturen oder hohe Luftfeuchtigkeit verändern. Darüber hinaus treten in einigen Fällen Fertigungsfehler auf, die z.B. zu objektivinternen Materialspannungen und damit zum Bruch von Linsen führen können.

67 Vgl. Schröder; Treiber 2007, S. 86.

68 Vgl. Meschede 2008, S. 174.

69 Vgl. Schmidt 2009, S. 384.

70 Vgl. Schröder; Treiber 2007, S. 86.

71 Vgl. Schröder; Treiber 2007, S. 86.

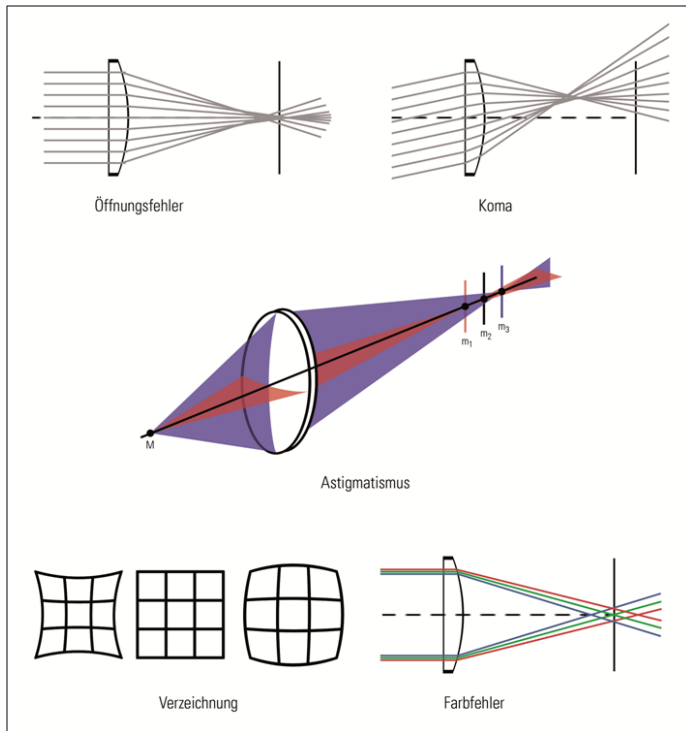


Abbildung 8: Abbildungsfehler

2.1.8.2 Bildfehler bei elektronischen Sensoren

- **Rauschen:** Dieser Bildfehler äußert sich durch unregelmäßige Flecken im Bild, die nicht den tatsächlichen Farben entsprechen, die eigentlich an den betreffenden Bildpunkten sichtbar sein müssten. Rauschen bei Bildsensoren entsteht in erster Linie durch die Grundspannung, die zum Betrieb des Sensors nötig ist. Diese verursacht ähnlich wie beim Einfall von Lichtstrahlen die Wanderung von Elektronen, daher verfälscht sich die Ladung bei einigen Photodioden auf dem Pixelraster des Chips⁷². Das Rauschen ist ein wichtiger Faktor hinsichtlich der Dynamik, denn diese bestimmt den Aussteuerbereich, dessen Obergrenze durch die Maximalladung eines einzelnen Pixels und dessen Untergrenze durch das Rauschen bestimmt ist. Die Dynamik ist daher ausschlaggebend für den durch den Chip erfassbaren Kontrastbereich⁷³. Über

⁷² Vgl. Schmidt 2009, S. 372.

⁷³ Vgl. Schmidt 2009, S. 372.

den Signal-Rauschabstand ergibt sich ein Maß für den Rauschanteil eines Signales⁷⁴. Beim Lichteinfall auf den Bildsensor entsteht wiederum zusätzliches Rauschen dadurch, dass nicht auf alle Pixel die identische Menge an Photonen auftrifft und so unterschiedliche Ladungen entstehen. Weitere Rauschquellen ergeben sich während des laufenden Betriebs aus der Entstehung von Wärme und die dem Sensor nachgeschalteten elektronischen Bauteile, z.B. Verstärker⁷⁵. Die Größe und der gegenseitige Abstand der einzelnen Pixel auf einem Chip sind bedeutsam für das Rauschen, größere Pixelflächen und Abstände bedeuten weniger Rauschen⁷⁶. Bei kleiner werdenden Flächen und Abständen können weniger in den Lichtstrahlen enthaltene Photonen auf einen einzelnen Pixel fallen, wodurch das Rauschen stärker wird.

- **Blooming:** Dieser Fehler tritt auf, wenn auf einen bestimmten Bereich des Chips Licht mit einer hohen Intensität fällt. In den dortigen Dioden bildet sich dann so viel Ladung, dass die Aufnahmekapazität überschritten wird. Dadurch geht ein Teil der Ladungsmenge auf benachbarte Pixel über, wodurch deren Ladungen erhöht werden und nicht mehr den eigentlich herrschenden Ladungen entsprechen⁷⁷. Somit erscheinen die benachbarten Flächen eines hellen Bildbereichs zusätzlich aufgehellte. Bei CMOS-Chips tritt dieser Fehler weniger stark auf, da überschüssige Ladung bei einem Pixel durch einen eigenen Transistor abgeführt werden kann⁷⁸.

- **Smear:** Dieser Bildfehler wurde bereits im Zusammenhang mit den Ausleseverfahren von CCD-Chips erläutert (siehe Abschnitt 2.1.2 Bildfeld). Verursacht wird der Fehler durch den Lichteinfall auf benachbarte Lichtdioden während des Ladungstransportes⁷⁹. Besonders bei hellen Stellen in einer dunklen Umgebung kann der Fehler deutlich sichtbar werden. Auch dieser Fehler tritt bei CMOS-Chips nicht auf, da die Ladungen vor dem Auslesen nicht verschoben werden.

- **Moiré:** Moirémuster treten dann auf, wenn sich regelmäßige Strukturen in einem bestimmten Winkel überlagern⁸⁰. Ein Beispiel hierfür sind zwei gleichmäßige Raster, die übereinander liegen und gegeneinander gedreht werden. Bei einem elektronischen Bildsensor entsteht ein Moiré durch die Interferenz zwischen einer Motivstruktur und dem Abtaststrat der Elemente des Bildsensors. Besonders deutlich tritt der Effekt bei feinen, detaillierten und gleichmäßigen Motivstrukturen auf wie z.B. bei Dachziegeln in einer Panoramaansicht oder

74 Vgl. Poynton 2003, S. 19.

75 Vgl. Schröder; Treiber 2007, S. 156.

76 Vgl. Böhringer; Bühler; Schlaich 2008, Bd. 2, S. 287.

77 Vgl. Böhringer; Bühler; Schlaich 2008, Bd. 2, S. 287.

78 Vgl. Litwiller 2001.

79 Vgl. Mahler 2005, S. 496.

80 Vgl. Böhringer; Bühler; Schlaich 2008, Bd. 2, S. 287f.

den feinen Streifen eines Hemdes.

- **Kompressionsartefakte:** Die Kodierung erfolgt über sog. Kompressionsverfahren. Bei verlustbehafteten Kompressionsverfahren, die sich auf mögliche Irrelevanzen der menschlichen visuellen Wahrnehmung abstützen, kann es vorkommen, dass der Informationsverlust doch sichtbar wird. Wenn z.B. benachbarte Bildpixel mit unterschiedlichen Farben zu einer gemeinsamen Fläche zusammengefasst werden, ist diese Blockbildung ab einer gewissen Größe sichtbar⁸¹. Können die Auswirkungen eines Kompressionsverfahrens wahrgenommen werden, so werden diese als Kompressionsartefakte bezeichnet. Die Ausprägungen solcher Artefakte sind insbesondere für die weitere Bildverarbeitung (computer vision) unerwünscht bzw. limitierend z.B. Farbverfälschungen⁸², Verlust von örtlicher Auflösung und Kontrast (Schärfe), Blockartefakte⁸³, unerwünschte Konturen, u.v.m.

2.2 Zusätzliche Kameraparameter

Neben den unmittelbar bildbestimmenden Parametern sind beim Einsatz von Kameras weitere geometrische Aspekte von Bedeutung. Obwohl sich diese Gesichtspunkte nicht direkt auf die Art und Weise auswirken, wie die Abbildung der Szene entsteht, so müssen sie dennoch beachtet werden – insbesondere, wenn es z.B. um die nachträgliche Rekonstruktion der räumlichen Tiefe einer Szene aus stereoskopischen Bildpaaren geht.

2.2.1 Position und Orientierung

Hier geht es um die Neigung der Kamera hinsichtlich der drei Raumachsen, d.h. das Rotationsverhalten sowie die Position der Kamera im Raum und relativ zur Szene. Es existieren in der Filmproduktion verschiedene Hilfsmittel zur Befestigung von Kameras, die unter dem Begriff Support zusammengefasst werden⁸⁴. Zu nennen sind hier verschiedene Stative, Bühnen, stationäre und mobile Kräne, Kamerawägen auf Rädern oder Schienen sowie Seilsysteme, bei denen Kameras an Seilbahnen entlang laufen. Weiterhin fallen in diesen Bereich sog. Steadycams, d.h. tragbare Halterungssysteme, die bei großer Bewegungsfreiheit für den Kameramann gleichzeitig verwacklungsarme Bilder ermöglichen. Auch der Einsatz von einem Fahrzeug oder einem anderen Fortbewegungsmittel für Land, Wasser oder Luft spielt in der Praxis oft eine Rolle. Je nach Größe und Gewicht der Kamera ist die Aufnahme natürlich auch ohne spezielle Befestigungen aus der Hand oder von der Schulter aus möglich. Abhängig von der Supportart weist eine Kamera zu jedem Zeitpunkt bestimmte Rotations- und Positionsparameter bezüglich der drei räumlichen Achsen und den Bezugspunkten

⁸¹ Vgl. Poynton 2003, S. 457.

⁸² Vgl. Poynton 2003, S. 515.

⁸³ Vgl. Poynton 2003, S. 471.

⁸⁴ Vgl. Schmidt 2009, S. 444.

der aufzunehmenden Szene auf. Die Bezugspunkte sind in diesem Fall alle Objektpunkte, die auf der Bildebene als Bildpunkte abgebildet werden. Die Raumtiefe in der Abbildung einer Szene steht daher immer im Zusammenhang mit der Position der Kamera relativ zur Szene. Diese zunächst scheinbar triviale Tatsache spielt für die weiteren Betrachtungen eine große Rolle.

2.2.2 Parameterveränderung bei laufender Aufnahme

Bei Kameras, die Bildsequenzen aufzeichnen, spielt der Faktor der Zeit eine große Rolle. Einige Kameraparameter können während der Aufnahme nicht verändert werden, z.B. die Flächengröße eines einzelnen Pixels auf einem Sensor, die Gesamtzahl der Pixel oder die Position des Bildfeldes. Ein Teil der Kameraparameter jedoch erlaubt eine Veränderung während der laufenden Aufnahme. Daher muss vor Aufnahmebeginn immer geklärt sein, ob und wie sich welche Parameter verändern sollen und wer dafür zuständig ist. Dabei ist es wichtig zu wissen, welche Parameter wie miteinander zusammenhängen. Wird eine Kamera beispielsweise bewegt, so bringt dies eine Veränderung im o.g. Verhältnis zwischen Szene und Kamera mit sich. Zu den Kameraparametern, die während der laufenden Aufnahme verändert werden können, zählen:

- Brennweite (nur bei Zoom-Objektiven) und damit Bildwinkel
- Durchmesser der Blendenöffnung
- Fokus
- Einbringung verschiedener Filter (abhängig vom beabsichtigten Zweck)
- Position und Rotation

Nur der Vollständigkeit wegen seien noch die technisch machbaren (aber in der Filmpraxis eher selten genutzten) Veränderungen der Bildaufnahmerate, der Verschlusszeit (shutter) und der Verstärkung und der Kamerakennlinie (gain) erwähnt.

2.2.3 Mathematische Repräsentation

Im weiteren Verlauf wird sich zeigen, dass Situationen existieren, in denen eine Kamera über einen einzelnen Punkt definiert wird, um in einen mathematischen Kontext gefasst werden zu können. Dieser Punkt stellt dann den Bezug für sämtliche Berechnungen bezüglich Translations- oder Rotationsänderungen der Kamera dar. Dies soll anhand eines einfachen Beispiels verdeutlicht werden. Eine bekannte Bildsequenz in der Filmproduktion stellt der horizontale Schwenk dar (panorama, kurz: pan). Dabei befindet sich eine Kamera auf einer Befestigung, z.B. auf einem Stativ und wird in horizontaler Richtung gedreht. Das Zentrum der Rotationsbewegung und somit der Bezugspunkt befindet sich hier im Drehpunkt des Stativkopfes. Im Hinblick auf den Prozess der optischen Abbildung müsste eine Kamera über ihr optisches Zentrum definiert werden, also den Brennpunkt des Objektivs C ⁸⁵. Wird dieser Punkt als

85 Vgl. Dobbert 2005, S. 118.

Bezugspunkt für Rotationen und Bewegungen herangezogen, so ergeben sich daraus vollkommen andere Bildveränderungen im Vergleich zu einem Bezugspunkt, der sich wie z.B. im Falle des Horizontalschwenks im Inneren des Stativkopfes befindet.

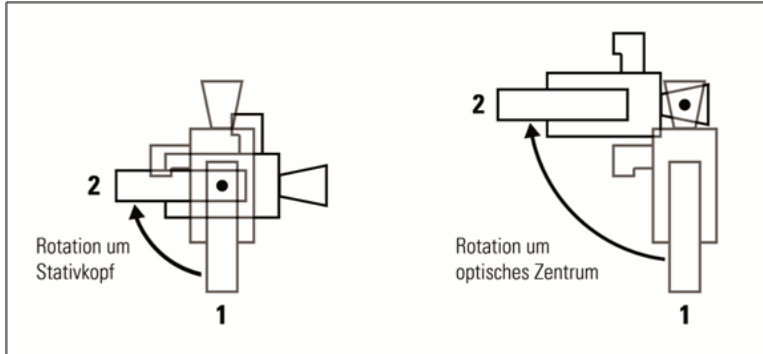


Abbildung 9: Kamerarotation

Daraus wird deutlich, dass eindeutig geklärt sein muss, über welchen Punkt eine Kamera in einem mathematischen Modell repräsentiert wird. Ist dies nicht gewährleistet, so ergeben sich daraus z.B. bei der Rekonstruktion von Bewegungen oder bei der Erzeugung von Tiefenkarten verfälschte Berechnungsergebnisse. In Computerprogrammen zur Modellierung und Animation von 3D-Objekten wie z.B. Autodesk Maya oder MAXON Cinema 4D können virtuelle Kameras erstellt werden. Diese werden über mehrere Parameter definiert^{86, 87} (die genauen Bezeichnungen variieren abhängig vom verwendeten Programm).

- Augpunkt Z : Dieser Punkt entspricht der Position der Kamera in Weltkoordinaten.
- Look-at-Punkt H : H spannt gemeinsam mit Z eine Gerade auf, die der Sichtungstrichtung der Kamera entspricht (Orientierung).
- Normale n : n ist ein Vektor, der sich durch die Gerade durch H und Z ergibt. Er steht orthogonal zur Projektionsebene, d.h. die Ebene, wo die Szene abgebildet wird.
- View-Up-Vektor: Dieser Vektor gibt eine Richtung an, die „oben“ entspricht.
- Art der Projektion: Dieser Parameter definiert, ob es sich um eine Zentral- oder Parallelprojektion handelt, d.h. ob eine perspektivische Verzerrung be-

⁸⁶ Vgl. Hofmeyer 2008, 122f.

⁸⁷ Vgl. Bender; Brill 2006, S. 36ff.

rücksichtigt wird (siehe nachfolgendes Kapitel).

- Grenzen des Sichtvolumens: Diese Grenzen entsprechen dem Bildwinkel und geben an, welcher Teil des Szenenbereichs in die Projektion eingeht. Das Sichtvolumen enthält zusätzlich zwei Entfernungswerte, die angeben, ab welcher Distanz zur Kamera ein Objekt als sichtbar definiert wird. Objekte, die außerhalb dieser Werte liegen, werden dann für die Projektion ignoriert.

Zusätzlich zu den o.g. Parametern können die übrigen Kameraeinstellungen (Brennweite, Blende, Größe der Bildebene, etc.) über verschiedene Menüs angegeben werden.

2.3 Kameraparameter eines Stereosystems

Bei Stereokamerasystemen erweitern sich die Parameter, die während der laufenden Aufnahme verändert werden können, um die beiden Aspekte Stereobasis und Konvergenzwinkel. Beide Komponenten wirken sich geometrisch auf unterschiedliche Weisen auf das Stereobildpaar aus – wahrnehmungspsychologisch und gestalterisch z.T. sehr erheblich. Aus diesem Grund sollten solche Parameteränderung bei einer stereoskopischen Aufnahme immer vorab hinsichtlich folgender Kriterien beurteilt werden:

1. Technische Umsetzbarkeit: Wie bereits erläutert, müssen Änderungen bei der Stereobasis und dem Konvergenzwinkel absolut synchron erfolgen, um die Stereowirkung aufrecht zu erhalten. Hierfür wird ein entsprechend leistungsfähiges Kamera-Rig benötigt. Verschiedene Systeme ermöglichen eine Parameteränderung lediglich bis zu einem gewissen Grad. Ein Beispiel stellt das 3D Live Rig der Firma Stereoscopic Technologies GmbH dar, welches eine minimale (maximale) Stereobasis von -10 mm (160 mm) erlaubt⁸⁸.
2. Sehempfinden: Handelt es sich bei der geplanten Änderung um einen Vorgang, zu welchem auch das visuelle System des Menschen fähig ist? Wenn dies nicht der Fall ist, besteht die Gefahr, dass die Änderung beim Betrachter Irritationen verursacht. Es ist zu beachten, dass dieses Kriterium lediglich für bestimmte Fälle gilt. Beispielsweise wird für Landschaftsaufnahmen die Stereobasis oft um das Vielfache des menschlichen Augenabstandes vergrößert, um ausreichend Tiefeninformation in die Stereoansichten einfließen zu lassen.
3. Szenenstruktur: Sind die beabsichtigten Änderungen hinsichtlich der Tiefenstruktur der aufzunehmenden Szene und der relativen Kameraposition sinnvoll? Soll beispielsweise die Stereobasis bei einer Nahaufnahme von 60 mm auf 200 mm vergrößert werden, so resultieren daraus zwei derartig unter-

88 Vgl. Stereoscopic Technologies GmbH (Hg.) 2010.

schiedliche Perspektiven, dass der Stereoeindruck für den Betrachter zunichte gemacht wird.

4. Dramaturgie: Das wichtigste Kriterium, nach dem eine Stereo-Parameteränderung beurteilt werden sollte, betrifft die dramaturgische Wirkung, die damit erzielt werden möchte. Die Änderung darf nicht zum reinen Selbstzweck verkommen.

Wenn für ein Stereokamerasystem mathematische Betrachtungen erfolgen sollen, ist es ebenso wie für eine einzelne Kamera notwendig, gewisse Hilfsgrößen zu definieren, die als Referenz für Berechnungen dienen. Als wichtige Größen sind z.B. die beiden optischen Zentren der Kameras C_1 und C_2 sowie die Stereobasis SB festzuhalten, welche die beiden Punkte verbindet. Für die Bestimmung des Konvergenzwinkels können mehrere Ansätze (siehe Abb. 10) genutzt werden ⁸⁹.

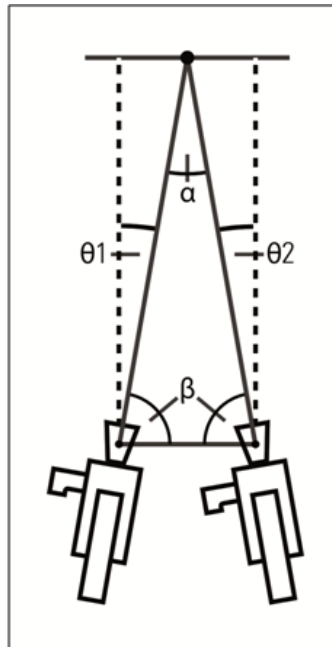


Abbildung 10: Konvergenzwinkel

Nach Lipton werden i. A. als Konvergenzwinkel bei stereoskopischen Filmaufnah-

⁸⁹ Vgl. Lipton 1982, S. 97.

men die Winkel θ_1 und θ_2 herangezogen⁹⁰. Unabhängig von der Messreferenz sollten die Winkel für beide Kameras gleich groß sein. Sind die Kameras nicht im gleichen Winkel ausgerichtet, besteht die Möglichkeit, die Bilder beider Ansichten über eine virtuelle Drehung – die Rektifikation – aneinander anzupassen (siehe nachfolgendes Kapitel). Generell jedoch sollten die Kameras so exakt wie möglich ausgerichtet sein und die Rektifikation nur dazu genutzt werden, diejenigen Fehler zu korrigieren, welche sich nicht vermeiden lassen. Sollen beide Kameras durch einen gemeinsamen Punkt repräsentiert werden, so kann dieser Punkt in der Mitte der Stereobasis-Strecke definiert werden. Dann muss jedoch beachtet werden, dass dieser Punkt allein keine Referenz bezüglich des Konvergenzwinkels bietet.

2.4 Zusammenfassung

Im Folgenden sind die wichtigsten Parameter aufgeführt, die bezüglich einer Bewegtbild-Kamera und einem digitalen Videosignal existieren und deren Angaben in der gesamten Produktionskette notwendig sind. In der Praxis liegen in den wenigsten Fällen alle Parameter vollständig und hinreichend exakt vor. Je nach Fall können bestimmte Angaben eine höhere Relevanz besitzen als andere.

Angaben über das Videoformat:

- Pixelraster (örtliche Auflösung)
- Zeilenaufbauprinzip
- Bildwechselfrequenz (zeitliche Auflösung)
- Abtastrate
- Quantisierung
- Chromasubsampling
- Kodierungsverfahren

Angaben zur Kamera:

- Brennweite (+ ggf. zeitliche Veränderung)
- Blendenzahl (+ ggf. zeitliche Veränderung)
- Bildwinkel (+ ggf. zeitliche Veränderung)
- Verschlusszeit, bzw. –winkel / Belichtungszeit
- Belichtungsmenge
- Zusätzliche optische Filter
- Objektiv: Typ, Modellbezeichnung, Hersteller
- Kamera: Typ, Modellbezeichnung, Hersteller
- Kamerakennlinie
- Befestigung, Aufbau, Support

90 Vgl. Lipton 1982, S. 98.

- Neigungswinkel für alle drei Raumachsen (+ ggf. zeitliche Veränderung)
- Entfernung des optischen Zentrums der Kamera zur Szene (+ ggf. zeitliche Veränderung)
- Höhe des optischen Zentrums der Kamera vom Boden (+ ggf. zeitliche Veränderung)
- Optische Verzerrungen, die durch das Objektiv verursacht werden. Die Verzerrungen entsprechen dem Bildfehler der Verzeichnung und werden üblicherweise über sog. Distortion Grids dokumentiert. In der Produktionspraxis handelt es sich dabei oft um ein Blatt Papier mit einem regelmäßigen Muster aus Quadraten mit bekannter Größe ⁹¹. Nach jedem Objektivwechsel wird das Blatt für einen kurzen Zeitraum aufgenommen, dann kann später der jeweilige Verzerrungsgrad ermittelt und die Ansicht entsprechend entzerrt werden.

Zusätzliche Angaben bei Filmkameras:

- Filmidentifikationsnummer
- Konfektionierung (Größe des Aufnahmeformats)
- Lichtempfindlichkeit (nach DIN-Norm oder ASA-Grad, oder über ISO-Nummer, welche die beiden ersten enthält ⁹²)
- Auflösungsvermögen
- Hersteller

Zusätzliche Angaben für Matchmoving:

- Abmessungsangaben für ein oder mehrere in der Szene sichtbare Referenzobjekte

Zusätzliche Angaben bei Stereosystemen:

- Stereobasis (+ ggf. zeitliche Veränderung)
- Konvergenzwinkel (+ ggf. zeitliche Veränderung)

⁹¹ Vgl. Dobbert 2005, S. 172.

⁹² Vgl. Ascher; Pincus 2007, S. 260.

3 Projektion und stereoskopische Projektion

3.1 Projektion

Der Raum, in welchem sich Menschen bewegen und handeln, besitzt drei Dimensionen. Diese drei Dimensionen werden i. d. R. mit Hilfe eines Koordinatensystems veranschaulicht, welches drei Achsen besitzt – Breite, Höhe und Tiefe. Die Achsen stehen orthogonal zueinander und schneiden sich in einem Ursprungspunkt. Dieses Konzept lässt sich dadurch erweitern, dass jeder Achse ausgehend von besagtem Ursprungspunkt zwei Richtungen zugewiesen werden (negativ und positiv). In der Mathematik erfolgt die Beschreibung des dreidimensionalen Raumes durch das Konzept des sog. euklidischen Raumes⁹³. Die Grundelemente des euklidischen Raumes sind u.a. Punkte, Linien, Geraden und Ebenen.

Bilder, die durch digitale oder analoge Fotografie entstehen, zeigen einen bestimmten Ausschnitt des dreidimensionalen Raumes in einer Darstellungsform, die lediglich zwei Dimensionen besitzt. Diese zweidimensionale Repräsentation des dreidimensionalen Raumes wird Projektion oder Abbildung genannt⁹⁴. Die Fläche, auf der das projizierte Bild entsteht, wird als Projektionsebene bezeichnet. Bei fotografischen Bildern vollziehen sich Projektionsvorgänge sowohl bei der Aufnahme als auch bei der Wiedergabe. In den meisten Fällen wird eine Projektion begleitet von Skalierungsprozessen, d.h. der Szenenausschnitt und sein Inhalt wird proportional vergrößert oder verkleinert.

Abbildungen, die über die sog. Parallelprojektion erzeugt werden, enthalten keine Informationen über die Tiefe des dreidimensionalen Raumes⁹⁵ in dem Sinne, dass Objekte mit gleicher Größe auch bei unterschiedlichen Entfernungen zur Projektionsebene gleich groß abgebildet werden. Der Faktor der Perspektive wird vernachlässigt, wodurch es beispielsweise möglich ist, anhand der projizierten Abbildung Boden- oder Gebäudeflächen korrekt zu vermessen. Die Parallelprojektion findet u.a. Verwendung in 3D-CAD- Systemen (Computer Aided Design), in der Architektur oder in militärischen Bereichen⁹⁶.

Die Projektionsart, die den Faktor der perspektivischen Verzerrung berücksichtigt, wird als Zentralprojektion oder perspektivische Projektion bezeichnet⁹⁷. Die Zentralprojektion entspricht der menschlichen Wahrnehmung und nach ihren Eigenschaften entstehen auch die meisten Abbildungen bei Aufnahmen mit Kameras⁹⁸. In speziellen

93 Vgl. Faugeras; Loung 2001, S. 4f.

94 Vgl. Dobbert 2005, S. 4.

95 Vgl. Bender; Brill 2006, S. 31.

96 Vgl. Faugeras; Loung 2001, S. 216.

97 Vgl. Faugeras; Loung 2001, S. 5.

98 Vgl. Steinmüller 2008, S. 204.

Fällen können Abbildungen erzeugt werden, die sich an die Parallelprojektion annähern. Der Übergang von der Zentralprojektion zu einer „Beinahe-Parallelprojektion“ oder umgekehrt wird von Kameraleuten als Dolly-Zoom oder Vertigo-Effekt bezeichnet^{99, 100}.

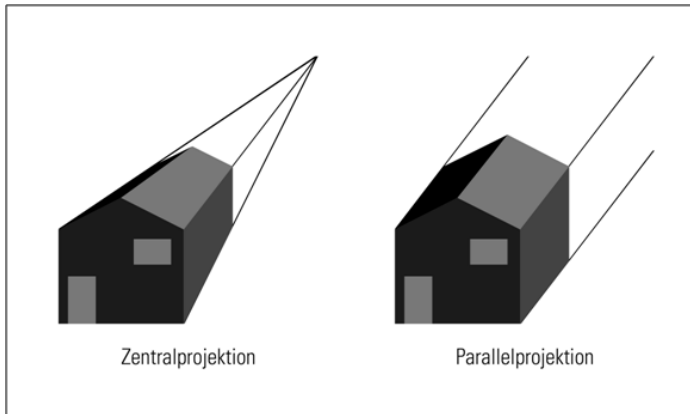


Abbildung 11: Zentral- und Parallelprojektion

Eine wesentliche Eigenschaft der Zentralprojektion besteht darin, dass sich untereinander parallele Linien des dreidimensionalen Raumes in Fluchtpunkten schneiden (mit Ausnahme von vertikal verlaufenden parallelen Linien)¹⁰¹. Existieren mehrere dieser Fluchtpunkte, so liegen alle auf einer horizontal verlaufenden Linie, welche entsprechend als Horizont (vanishing line) bezeichnet wird. Dies führt dazu, dass die Winkel-, Längen- und Größenverhältnisse von Szenenobjekten perspektivisch verzerrt abgebildet werden¹⁰². Zwei Objekte mit eigentlich identischer Größe weisen daher bei unterschiedlicher Entfernung zur Projektionsebene verschiedene Größenverhältnisse auf.

3.2 Projektion bei einer einzelnen Kamera

Der Abbildungsprozess der perspektivischen Projektion lässt sich mit Hilfe eines Lochkamera-Modells anschaulich darstellen. Alle Szenenpunkte des dreidimensionalen Raumes werden über ein optisches Zentrum auf der zweidimensionalen Bildebene (die der Projektionsebene entspricht) abgebildet¹⁰³. Das optische Zentrum

99 Vgl. Schreer 2005, S. 53.

100 Vgl. Ascher; Pincus 2007, S. 356.

101 Vgl. Schreer 2005, S. 11.

102 Vgl. Bender; Brill 2006, S. 32.

103 Vgl. Hartley; Zisserman 2003, S. 153.

entspricht dem Brennpunkt C , welcher auf der Brennebene liegt. Bezüglich der Projektionsart befindet sich das optische Zentrum bei der Parallelprojektion in unendlicher Entfernung und bei der Zentralprojektion in endlicher¹⁰⁴. Im Falle der Lochkamera erfährt die Abbildung eine Punktspiegelung am Brennpunkt, so dass das Bild auf dem Kopf stehend projiziert wird. Zur Vereinfachung wird für den weiteren Verlauf dieser Arbeit eine Darstellung gewählt, bei der die Bildebene vor der Brennebene liegt. Bis auf eine Spiegelung der x - und y -Koordinaten wirkt sich dies nicht auf die mathematischen Zusammenhänge aus¹⁰⁵.

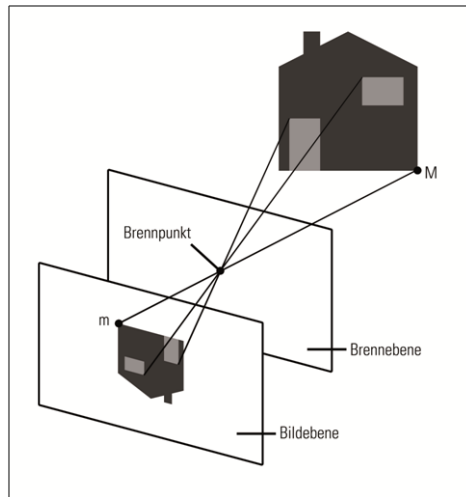


Abbildung 12: Lochkameramodell

Ein Szenenpunkt $M_w = (X_w, Y_w, Z_w)^T$ eines dreidimensionalen Objekts wird auf der Bildebene I im 2D-Punkt $m = (u, v)^T$ abgebildet. Der Mittelpunkt der Ebene I wird als Kamerahauptpunkt c bezeichnet. Bezogen auf die technischen Komponenten einer realen Kamera stellt I einen oder mehrere Bildsensoren dar¹⁰⁶. Die optische Achse der Kamera entspricht einer Geraden durch den Brennpunkt C und den Kamerahauptpunkt c . Der Abstand zwischen den beiden Punkten entspricht der Brennweite f , welche identisch ist mit dem Abstand zwischen der Hauptebene des optischen Systems und dem Brennpunkt (siehe vorheriges Kapitel). Es existieren bei diesem Modell insgesamt vier unterschiedliche Koordinatensysteme:

- Weltkoordinatensystem: Koordinatensystem der Szene, drei Achsen, metrische Einheiten

¹⁰⁴ Vgl. Faugeras; Loung 2001, S. 199.

¹⁰⁵ Vgl. Schreer 2005, S. 41.

¹⁰⁶ Vgl. Ascher; Pincus 2007, S. 141.

- Kamerakoordinatensystem: Koordinatensystem bezüglich der räumlichen Orientierung der Kamera, drei Achsen, metrische Einheiten
- Sensorkoordinatensystem: zweites Koordinatensystem der Kamera, Fläche des Bildsensors, zwei Achsen, metrische Einheiten
- Bildkoordinatensystem: Koordinatensystem des Bildes, zwei Achsen, diskrete Einheiten

Das Sensorkoordinatensystem und das Bildkoordinatensystem liegen in ein und derselben Ebene, unterscheiden sich jedoch in der Position des Ursprungspunktes und den Maßeinheiten. Bei besagter Ebene handelt es sich um die bereits erwähnte Bildebene *I*. Nach Schreer lässt sich die optische Abbildung in drei Prozesse unterteilen, welche als externe, interne und perspektivische Transformation bezeichnet werden¹⁰⁷. Bevor eine detaillierte Beschreibung dieser Vorgänge folgt, muss kurz auf das Konzept der homogenen Koordinaten eingegangen werden. Mathematisch betrachtet handelt es sich bei der Zentralprojektion um eine sog. projektive (kollineare) Abbildung. Die Projektionsbeziehung zwischen dem Szenenraum und dem Bildraum wird als räumliche Ähnlichkeitstransformation bezeichnet¹⁰⁸. Da die homogenen Koordinaten die Darstellung verschiedener Transformationen in einer einheitlichen Matrix ermöglichen, spielen sie in der projektiven Geometrie eine zentrale Rolle. Zusätzlich können komplexe Transformationen durch Matrixmultiplikation einfacher Transformationen realisiert werden¹⁰⁹.

3.2.1 Homogene Koordinaten

Homogene Koordinaten geben einen Punkt in Vektorform an¹¹⁰. Im Falle des Punktes $M_w = (X_w, Y_w, Z_w)^T$ erfolgt die Darstellung in homogenen Koordinaten als $\tilde{M}_w = (hX_w, hY_w, hZ_w, h)^T$. Es ist zu erkennen, dass der Vektor der Dimension n um eine Dimension erweitert wird, wobei die zusätzliche Komponente h einen Skalierungsfaktor darstellt. Der Vektor ist somit in seiner Richtung definiert, jedoch nicht in seiner Länge. Um die ursprünglichen metrischen Koordinaten von M_w zu erhalten, müssen die ersten drei Werte durch den vierten dividiert werden¹¹¹.

3.2.1 Externe Transformation

Als Ursprung für das Koordinatensystem der Kamera dient das bereits erwähnte optische Zentrum der Kamera, also der Brennpunkt *C*. Dieser befindet sich im Objektiv, seine genaue Position ist nicht exakt messbar. Um eine eindeutige mathematische Relation zwischen der Szene und der Kamera herzustellen, wird eine sog. euklidische

¹⁰⁷ Vgl. Schreer 2005, S. 42.

¹⁰⁸ Vgl. Wikipedia: Zentralprojektion.

¹⁰⁹ Vgl. Steinmüller 2008, S. 212.

¹¹⁰ Vgl. Hartley; Zisserman 2003, S. 27.

¹¹¹ Vgl. Steinmüller 2008, S. 212.

Transformation durchgeführt, bei der über eine Translation \mathbf{t} und eine Rotation \mathbf{R} das Koordinatensystem der Szene mit dem Ursprung O in das Koordinatensystem der Kamera mit dem Ursprung C überführt wird ¹¹². Die Beziehung zwischen ein und demselben Punkt im dreidimensionalen Koordinatensystem der Szene (M_w) und im dreidimensionalen Koordinatensystem der Kamera (M_c) lässt sich demnach folgendermaßen beschreiben:

$$M_c = \mathbf{R} M_w + \mathbf{t} \quad \text{mit } \mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \text{ und } \mathbf{t} = \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix}$$

Formel 8

Die 3×3 -Matrix \mathbf{R} beschreibt die Orientierung der Kamera im Raum ¹¹³. Der 3×1 -Vektor \mathbf{t} repräsentiert die Verschiebung des Ursprungs des Weltkoordinatensystems ¹¹⁴. Sowohl \mathbf{R} als auch \mathbf{t} besitzen je drei Freiheitsgrade ¹¹⁵. Unter Nutzung der Darstellungsform der homogenen Koordinaten lassen sich \mathbf{R} und \mathbf{t} in einer gemeinsamen Matrix \mathbf{D} darstellen:

$$\tilde{M}_c = \mathbf{D} \tilde{M}_w \quad \text{mit } \mathbf{D} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}_3^T & 1 \end{bmatrix} \text{ und } \mathbf{0}_3 = [0,0,0]^T$$

Formel 9

Die Rotation und die Translation bilden die sog. externen Kameraparameter. Da sie in der Matrix \mathbf{D} zusammengefasst werden, bezeichnet man \mathbf{D} entsprechend als extrinsische Matrix. Nach der externen Transformation verläuft die Z-Achse des Weltkoordinatensystems entlang der optischen Achse der Kamera und schneidet die Bildebene I in ihrem Mittelpunkt, dem Kamerahauptpunkt c . Zu beachten ist, dass die Matrixmultiplikation nicht kommutativ ist, daher ist die Rotation ebenfalls nicht kommutativ. Die Reihenfolge, in der die Rotationen durchgeführt werden, muss aus diesem Grund eingehalten werden ¹¹⁶.

¹¹² Vgl. Jähne 2005, S. 200.

¹¹³ Vgl. Hartley; Zisserman 2003, S. 156.

¹¹⁴ Vgl. Steinmüller 2008, S. 212.

¹¹⁵ Vgl. Faugeras; Loung 2001, S. 208.

¹¹⁶ Vgl. Jähne 2005, S. 202.

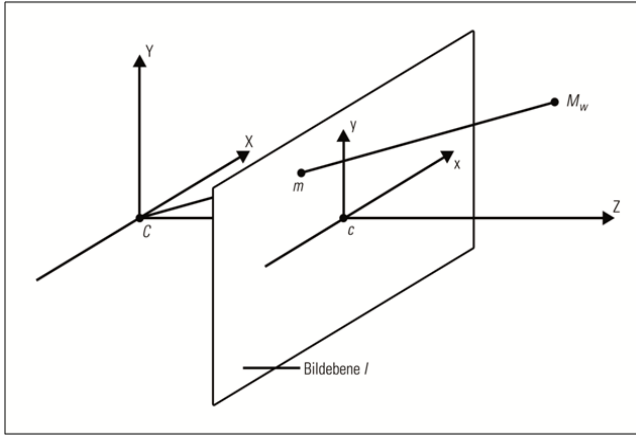


Abbildung 13: Kamerakoordinatensystem

3.2.3 Perspektivische Transformation

Der Punkt $M_c = (X_c, Y_c, Z_c)^T$ wird nun überführt vom dreidimensionalen Kamerakoordinatensystem in den Punkt $\mathbf{m}' = (x, y)^T$ des zweidimensionalen Koordinatensystems der Sensorfläche. Die Gleichung für die perspektivische Transformation lautet nach Schreer¹¹⁷:

$$\frac{x}{X_c} = \frac{y}{Y_c} = \frac{f}{Z_c}$$

Formel 10

Da die Brennweite f die Entfernung zwischen Brennpunkt C und Bildebene I bezeichnet, repräsentiert sie ebenfalls die Distanz zwischen C und der Sensorfläche. Die Beziehung zwischen M_c und \mathbf{m}' in Matrixform lautet dann¹¹⁸:

$$\begin{bmatrix} U \\ V \\ S \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} * \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \quad \text{mit } x = \frac{U}{S}, y = \frac{V}{S}, \text{ für } S \neq 0$$

Formel 11

¹¹⁷ Vgl. Schreer 2005, S. 43.

¹¹⁸ Vgl. Schreer 2005, S. 43.

Die Darstellung in homogenen Koordinaten lautet:

$$s\tilde{\mathbf{m}}' = \mathbf{P}'\tilde{\mathbf{M}}_c \quad \text{mit } \mathbf{P}' = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \text{ und } s = S$$

Formel 12

\mathbf{P}' wird als perspektivische Transformationsmatrix bezeichnet. Mit Hilfe der homogenen Koordinaten ist es möglich, die externe und die perspektivische Transformation des dreidimensionalen Raumpunktes $\tilde{\mathbf{M}}_w$ in den zweidimensionalen Sensorflächenpunkt \mathbf{m}' in einer einheitlichen Gleichung zusammenzufassen:

$$s\tilde{\mathbf{m}}' = \mathbf{P}' \mathbf{D} \tilde{\mathbf{M}}_w$$

Formel 13

3.2.4 Interne Transformation

Der letzte Schritt der perspektivischen Projektion besteht darin, die zu diesem Zeitpunkt noch metrischen Sensorkoordinatenwerte in diskrete Bildkoordinatenwerte zu überführen¹¹⁹. Für diesen Prozess sind zwei Operationen notwendig. Beim ersten Vorgang handelt es sich um eine horizontale und vertikale Skalierung k_u, k_v , mit deren Hilfe die metrischen Werte in diskrete Pixelwerte umgerechnet werden. Über die Skalierung kann auch der Fall berücksichtigt werden, wenn die lichtempfindlichen Dioden des Bildsensors rechteckige anstatt quadratische Formen aufweisen¹²⁰. Der zweite Vorgang beinhaltet eine Verschiebung des Ursprungspunktes c ¹²¹. Zu Beginn dieses Textes wurde erläutert, dass der Ursprung eines Bildkoordinatensystems entsprechend der Matrixnotation in der linken oberen Ecke des Bildes liegt. Der Ursprung des Sensorkoordinatensystems liegt jedoch im Kamerahauptpunkt c , also im Zentrum der Sensorfläche. Die Verschiebung des Ursprungs erfolgt durch eine zweidimensionale Translation (u_0, v_0) ¹²². Die entsprechende Transformation lautet dann:

$$\tilde{\mathbf{m}} = \mathbf{H} \tilde{\mathbf{m}}' \quad \text{mit } \mathbf{H} = \begin{bmatrix} k_u & 0 & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

Formel 14

¹¹⁹ Vgl. Schreer 2005, S. 21f.

¹²⁰ Vgl. Faugeras; Loung 2001, S. 207f.

¹²¹ Vgl. Wöhler 2009, S. 5f.

¹²² Vgl. Hartley; Zisserman 2003, S. 155.

Die (1,2)-Komponente von H lässt sich als Scherungsparameter berücksichtigen für den Fall, dass die Achsen des Bildsensors eine schiefsymmetrische Ausrichtung aufweisen ¹²³. Der überwiegende Anteil der Bildsensor-Chips für Kameras, die in der Filmindustrie zum Einsatz kommen, wird jedoch mit hoher Genauigkeit produziert, so dass dieser Parameter nach einer kurzen Überprüfung in den meisten Fällen vernachlässigt, bzw. gleich Null gesetzt werden kann. Die beiden Matrizen H und P' lassen sich aufgrund ihrer Dimensionen zusammenfassen, wodurch sich die perspektivische und die interne Transformation in einer gemeinsamen Matrix darstellen lassen ¹²⁴:

$$s\tilde{m}' = P_{neu} \tilde{M}_c \quad \text{mit } P_{neu} = H P' = \begin{bmatrix} f k_u & 0 & u_0 & 0 \\ 0 & f k_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Formel 15

Wird die Brennweite f , also der Abstand der Bildebene vom Ursprung des Kamerakoordinatensystems und die Distanz zwischen den Punkten C und c mit $f=1$ definiert, so handelt es sich um sog. normierte Koordinaten ¹²⁵. Mit dem geeigneten Faktor ist die Normierung für jede Kameramatrix zu bewerkstelligen ¹²⁶. Ist dies der Fall, so besitzt die Matrix der perspektivischen Transformation folgende Form:

$$P_{Norm} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Formel 16

Übertragen auf die Matrix, die sich aus H und P' zusammensetzt, ergibt sich ¹²⁷:

$$P_{neu} = A P_{Norm} \quad \text{mit } A = \begin{bmatrix} a_u & 0 & u_0 \\ 0 & a_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{für } a_u = f k_u \text{ und } a_v = f k_v$$

Formel 17

Damit enthält die Matrix A die intrinsischen Kameraparameter, bestehend aus den fünf Komponenten des Scherungsparameters, der horizontalen und vertikalen Skalierung über k_u und k_v sowie der zweidimensionalen Translation (u_0, v_0) ¹²⁸. Konse-

123 Vgl. Faugeras; Loung 2001, S. 208.

124 Vgl. Schreer 2005, S. 45.

125 Vgl. Schreer 2005, S. 45.

126 Vgl. Hartley; Zisserman 2003, S. 162.

127 Vgl. Schreer 2005, S. 46.

128 Vgl. Faugeras; Loung 2001, S. 210

quenterweise wird A daher als intrinsische Matrix bezeichnet. Es ergibt sich für die Darstellung der internen Transformation in homogenen Koordinaten:

$$\tilde{m} = A \tilde{m}'$$

Formel 18

3.2.5 Vollständiger Übergang von 3D-Szene zu 2D-Bild

Die gesamte Transformation vom Szenenpunkt M_w zum Bildpunkt m lässt sich anhand des bisher Dargelegten folgendermaßen definieren ¹²⁹:

$$s\tilde{m} = A \tilde{m}' = A * P_{Norm} \tilde{M}_c = A P_{Norm} * D \tilde{M}_w = A [R \ t] \tilde{M}_w = P \tilde{M}_w$$

Formel 19

$$\text{mit } P = A [R \ t] = \begin{bmatrix} a_u r_1^T + u_0 r_3^T & a_u t_x + u_0 t_z \\ a_v r_2^T + v_0 r_3^T & a_v t_y + v_0 t_z \\ r_3^T & t_z \end{bmatrix}$$

Formel 20

Insgesamt wird M_w also dreimal von einem Koordinatensystem in ein anderes überführt, bevor er zu m wird. Die 3×4 -Matrix P ist die allgemeine Form der perspektivischen Projektionsmatrix. Bis auf einen Skalierungsfaktor wird durch sie der Zusammenhang hergestellt zwischen einem dreidimensionalen Szenenpunkt und dessen Abbildung auf der Bildebene in Pixelkoordinaten. Wird – wie in der Praxis oft üblich – der Scherungsparameter vernachlässigt, so ergeben sich dadurch für P aufgrund ihrer Dimension 10 Freiheitsgrade. Diese entsprechen den Komponenten der externen Transformation (3 für die Rotation, 3 für die Translation) und der internen Transformation (2 für die Skalierung, 2 für die Verschiebung des Ursprungspunktes) ¹³⁰. Mit Hilfe der Matrixparameter lassen sich Gleichungen für u und v aufstellen, wobei die Angabe q_{mn} den jeweiligen Parameter aus P repräsentiert ¹³¹:

$$u = \frac{q_{11}X_w + q_{12}Y_w + q_{13}Z_w + q_{14}}{q_{31}X_w + q_{32}Y_w + q_{33}Z_w + q_{34}}$$

Formel 21

¹²⁹ Vgl. Schreer 2005, S. 46.

¹³⁰ Vgl. Faugeras; Loung 2001, S. 210.

¹³¹ Vgl. Schreer 2005, S. 46.

$$v = \frac{q_{21}X_w + q_{22}Y_w + q_{23}Z_w + q_{24}}{q_{31}X_w + q_{32}Y_w + q_{33}Z_w + q_{34}}$$

Formel 22

3.2.6 Linsenverzerrungen

Im vorherigen Kapitel wurde der Fehler der Verzeichnung erläutert. Dieser Fehler tritt sehr oft dann auf, wenn kurze Brennweiten zum Einsatz kommen. Mathematisch betrachtet handelt es sich dabei um eine nichtlineare Transformation von unverzerrten in verzerrte Koordinaten¹³². Die Verzerrung setzt sich zusammen aus radialen und tangentialen Anteilen, wobei hauptsächlich die Auswirkungen der radialen Verzerrung, die durch den Koeffizienten κ_i beschrieben wird, negativ auffallen¹³³. Für die Verzerrung lässt sich mit Hilfe der verzerrten (u_d, v_d) und unverzerrten Koordinaten (u, v) folgendes Modell aufstellen:

$$\begin{pmatrix} u_d \\ v_d \end{pmatrix} = L(r) \begin{pmatrix} u \\ v \end{pmatrix}$$

Formel 23

Dabei repräsentiert r die radiale Entfernung $\sqrt{u^2 + v^2}$ vom Verzerrungszentrum und $L(r)$ den Verzerrungsfaktor als Funktion von r . Um die Verzerrung zu korrigieren, ergeben sich für die Pixelkoordinaten folgende Formeln¹³⁴:

$$\hat{u} = u_c + L(r)(u - u_c)$$

Formel 24

$$\hat{v} = v_c + L(r)(v - v_c)$$

Formel 25

Die Elemente u und v stellen die gemessenen Koordinatenwerte dar und \hat{u} und \hat{v} sind die korrigierten Koordinaten. (u_c, v_c) beschreiben über $r^2 = (u - u_c)^2 + (v - v_c)^2$ das Zentrum der Verzerrung.

Die bis zu diesem Zeitpunkt beschriebenen Prozesse beziehen sich auf Aufnahmen mit einer einzelnen Kamera. Kommen Stereosysteme zum Einsatz, so erweitern sich

¹³² Vgl. Schreer 2005, S. 46.

¹³³ Vgl. Hartley; Zisserman 2003, S. 189.

¹³⁴ Vgl. Hartley; Zisserman 2003, S. 191.

die mathematischen Gesetzmäßigkeiten der perspektivischen Projektion um weitere Komponenten und mathematische Beziehungen. Diese Gesetze werden zusammengefasst in der sog. Epipolargeometrie. Mit der Epipolargeometrie können korrespondierende Punkte bestimmt werden, d.h. die jeweiligen Abbildungen ein und desselben Objektpunktes in den beiden verschiedenen Ansichten. Die Bestimmung von Punktkorrespondenzen ist z.B. notwendig zum Ermitteln der horizontalen Disparitäten, welche wiederum die Grundlage für die Erstellung einer Tiefenkarte bilden.

3.3 Projektion bei einem Stereosystem

3.3.1 Epipolargeometrie

Wie bereits erläutert, lassen sich Stereokameraanordnungen in die zwei Bereiche der achsparallelen und der konvergenten Systeme einteilen. Da die Anordnung mit zueinander gedrehten Kameras besonders in der industriellen Bildverarbeitung und der Computer Vision Einsatz findet, wird sie als Standardstereogeometrie bezeichnet¹³⁵. Viele der mathematischen Beziehungen besitzen im achsparallelen Fall eine einfachere Form. Aus diesem Grund lässt sich beispielsweise eine nachträgliche Korrektur der Disparität einfacher bewerkstelligen als bei Bildpaaren, die mit konvergenten Kameras aufgenommen wurden. Einer der Nachteile eines Systems mit parallelen Achsen besteht darin, dass eine Szene u.U. nicht vollständig erfasst werden kann, da die Kameras nicht auf einen gemeinsamen Punkt hin ausgerichtet sind¹³⁶. Hinsichtlich der Aufnahme von digitalen Bewegtbildern für Film und Video lässt sich sagen, dass mittlerweile der konvergente Ansatz in virtual sets¹³⁷ durchaus Verbreitung gefunden hat¹³⁸. Dabei werden die Kameras so ausgerichtet, dass sich der Schnittpunkt der optischen Achsen dort befindet, wohin die Aufmerksamkeit des Zuschauers gelenkt werden soll, ähnlich dem Einsatz der Tiefenunschärfe bei Aufnahmen mit einer Kamera¹³⁹.

3.2.1 Konvergentes System

3.3.2.1 Modellierung

Es existieren zwei optische Zentren bzw. Brennpunkte C_1 und C_2 sowie zwei Bildebenen I_1 und I_2 . Der dreidimensionale Szenenpunkt M_w wird auf die beiden Bildpunkte m_1 und m_2 abgebildet. Die beiden Bildebenen sind zueinander gedreht, wobei als Ausgangspunkt dieser Rotation die jeweiligen Brennpunkte dienen¹⁴⁰. Der Abstand zwischen je einem der Brennpunkte und seiner zugehörigen Bildebene wird über die

¹³⁵ Vgl. Steinmüller 2008, S. 222.

¹³⁶ Vgl. Schreer 2005, S. 65.

¹³⁷ In denen die potenziell starken retinalen Rivalitäten im Fernpunkt (= Grennscreen oder Blue-screen) keine wahrnehmungspsychologische Wirkung haben

¹³⁸ Vgl. Seymour 2008.

¹³⁹ Vgl. Cohen 2008.

¹⁴⁰ Vgl. Wöhler 2009, S. 10.

Brennweite f definiert und für beide Kameras als identisch festgelegt. Die Distanz zwischen den beiden Brennpunkten stellt die Stereobasis SB dar ¹⁴¹. Für die beiden Kamerakoordinatensysteme bedeutet dies, dass ähnlich wie bei der externen Transformation eine Rotation und eine Translation durchgeführt werden muss, um eine mathematische Beziehung zwischen beiden Koordinatensystemen herzustellen ¹⁴²:

$$M_{C_2} = R M_{C_1} + t$$

Formel 26

Dadurch, dass die beiden Kameras zueinander gedreht sind, werden beide Bildebenen I_1 und I_2 von der Linie der Stereobasis an je einem Punkt durchstoßen. Diese Punkte werden als Epipole e_1 und e_2 bezeichnet und repräsentieren die theoretische Abbildung des einen Brennpunkts auf der jeweils anderen Bildebene (C_1 auf I_2 , C_2 auf I_1) ¹⁴³. Gemeinsam mit dem Szenenpunkt M_w spannen die beiden Epipole die Epipolarebene Π auf. Dadurch befinden sich die beiden Bildpunkte m_1 und m_2 ebenfalls in der Epipolarebene. Mit dem Bildpunkt und dem Epipol, welche beide auf der gleichen Bildebene liegen, kann eine Gerade aufgespannt werden. Innerhalb den Grenzen eines Bildes wird diese Gerade als Epipolarlinie l bezeichnet ¹⁴⁴.

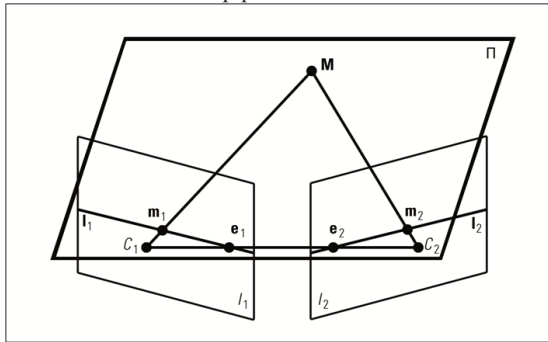


Abbildung 14: Epipolargeometrie, konvergent

Es wird deutlich, dass der Szenenpunkt M_w auf der Bildebene I_1 auch bei unterschiedlicher Entfernung immer an der gleichen Stelle abgebildet wird, solange er sich auf der Geraden befindet, die durch C_1 und m_1 verläuft. Auf der Bildebene I_2 ändert sich dabei die Stelle, an der M_w in m_2 abgebildet wird. Diese Positionsänderung erfolgt lediglich entlang der Epipolarlinie in I_2 . Das gleiche Prinzip gilt für eine Positionsänderung von M_w entlang der Geraden durch C_2 und m_2 . Für jeden weiteren Szenenpunkt

¹⁴¹ Vgl. Lipton 1982, S. 136.

¹⁴² Vgl. Schreier 2005, S. 68.

¹⁴³ Vgl. Faugeras; Loung 2001, S. 260.

¹⁴⁴ Vgl. Hartley; Zisserman 2003, S. 240.

entsteht ein weiteres Paar von Epipolarlinien. Sämtliche Epipolarlinien in einer Bildebene werden zusammengefasst unter dem Begriff Epipolarlinienbüschel. Alle Epipolarlinien eines Bildes schneiden sich im Epipol, wobei die Positionen der beiden Epipole unabhängig von der Position der Szenenpunkte sind¹⁴⁵. Wird für einen beliebigen Bildpunkt \mathbf{m}_1 der korrespondierende Bildpunkt \mathbf{m}_2 gesucht, so lässt sich die Suche auf den Bereich der zugehörigen Epipolarlinie eingrenzen¹⁴⁶. Anders als bei einem achsparallelen System verlaufen die Epipolarlinien beim konvergenten System lediglich horizontal¹⁴⁷. Wird daher eine Punktkorrespondenz für einen Bildpunkt gesucht, ist es notwendig, zuvor die korrekten Epipolarlinien zu ermitteln. Zu diesem Zweck müssen die mathematischen Beziehungen innerhalb des Systems definiert werden.

3.2.2.2 Mathematische Beziehungen in Kamerakoordinaten

Die Koordinatensysteme von Kamera 1 und Kamera 2 stehen über eine Rotation und eine Translation zueinander in Beziehung. Wird der Ursprung des Weltkoordinatensystems in den Ursprung des Koordinatensystems von Kamera 1 gelegt, so ergibt sich folgende mathematische Beziehung¹⁴⁸.

$$\begin{bmatrix} X_2 \\ Y_2 \\ Z_2 \end{bmatrix} = \mathbf{R} \begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \end{bmatrix} + \mathbf{t} \text{ mit } \mathbf{R} = \begin{bmatrix} \mathbf{r}_1^T \\ \mathbf{r}_2^T \\ \mathbf{r}_3^T \end{bmatrix}$$

Formel 27

Für die Bildpunkte von M_w ergeben sich die Sensorkoordinaten in Schreibweise der homogenen Koordinaten¹⁴⁹:

$$\tilde{\mathbf{m}}'_1 = \frac{M_{w1}}{Z_1} \text{ und } \tilde{\mathbf{m}}'_2 = \frac{M_{w2}}{Z_2}$$

Formel 28

Werden die Gleichungen kombiniert, ergibt sich für die Beziehung zwischen \mathbf{m}_1 und \mathbf{m}_2 eine Formel, die die beiden Unbekannten Z_1 und Z_2 aufweist¹⁵⁰:

$$\tilde{\mathbf{m}}'_2 = \frac{1}{Z_2} (Z_1 \mathbf{R} \tilde{\mathbf{m}}'_1 + \mathbf{t})$$

Formel 29

¹⁴⁵ Vgl. Hartley; Zisserman 2003, S. 241.

¹⁴⁶ Vgl. Steinmüller 2008, S. 228.

¹⁴⁷ Vgl. Faugeras; Loung 2001, S. 291.

¹⁴⁸ Vgl. Schreer 2005, S. 70.

¹⁴⁹ Vgl. Schreer 2005, S. 70.

¹⁵⁰ Vgl. Schreer 2005, S. 70.

Das Kreuzprodukt der Gleichung mit \mathbf{t} und das Skalarprodukt mit $\tilde{\mathbf{m}}_2'$ sowie die Eliminierung der beiden Unbekannten liefert folgende Beziehung¹⁵¹:

$$\mathbf{t} \times \tilde{\mathbf{m}}_2' = \frac{z_1}{z_2} \mathbf{t} \times \mathbf{R} \tilde{\mathbf{m}}_1' \text{ und } \tilde{\mathbf{m}}_2' * (\mathbf{t} \times \tilde{\mathbf{m}}_2') = \tilde{\mathbf{m}}_2'^T \mathbf{t} \times (\mathbf{R} \tilde{\mathbf{m}}_1') = 0$$

Formel 30

Wird eine Zuordnung eines dreidimensionalen Vektors zu einer antisymmetrischen Matrix entsprechend der folgenden Gleichung definiert, so lässt sich das Kreuzprodukt zweier Vektoren durch das Matrixprodukt-Vektor-Produkt einer 3×3-Matrix mit einem dreidimensionalen Vektor ausdrücken¹⁵².

$$\begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}_\times = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} \text{ mit } [\mathbf{t}]_\times = -[\mathbf{t}]_\times^T$$

Formel 30

Die Gleichung, die auf diese Weise zustande kommt, stellt die zentrale Gleichung für die mathematische Beziehung zwischen den beiden Abbildungen eines 3D-Punktes in zwei Kameras dar und wird als Epipolargleichung bezeichnet^{153, 154}:

$$\tilde{\mathbf{m}}_2'^T \mathbf{E} \tilde{\mathbf{m}}_1' = 0 \text{ mit } \mathbf{E} = [\mathbf{t}]_\times \mathbf{R}$$

Formel 32

Die Gleichung ist erfüllt, wenn die Bildpunkte in Kamera 1 und Kamera 2 korrespondierende Abbildungen eines dreidimensionalen Szenenpunktes sind. Es gilt jedoch nicht der Umkehrschluss, dass Abbildungen in beiden Kameras, welche die Epipolarbedingung erfüllen, auch korrespondierende Punkte sind. Der Grund dafür ist, dass alle Punkte auf einem optischen Strahl zwar die selbe Abbildung in einer Ansicht erzeugen, jedoch verschiedene Abbildungen in der anderen Ansicht¹⁵⁵. Die Matrix \mathbf{E} wird als Essential-Matrix bezeichnet. Sie enthält fünf unbekannte Parameter, zwei für die Verschiebungsrichtung zwischen den Zentren der Kamerakoordinatensysteme und drei für die relative Rotation¹⁵⁶. Die Essential-Matrix beschreibt nicht nur die

151 Vgl. Wöhler 2009, S. 12.

152 Vgl. Faugeras; Loung 2001, S. 12.

153 Vgl. Schreer 2005, S. 75.

154 Vgl. Hartley; Zisserman 2003, S. 153.

155 Vgl. Schreer 2005, S. 71.

156 Vgl. Faugeras; Loung 2001, S. 282.

euklidische Transformation von Kamera 1 nach Kamera 2 sondern auch gleichzeitig die Beziehung für Punkte in Sensorkoordinaten. Dies bedeutet, dass die intrinsischen Parameter bekannt sein müssen, um von den Bildkoordinaten in Pixel in Sensorkoordinaten in z.B. mm transformieren zu können ¹⁵⁷. In diesem Fall spricht man von kalibrierten Kameras.

3.3.2.3 Epipolarlinien und Epipol \tilde{e}_i'

1. Epipolarlinien: Eine Gerade in einer Ebene lässt sich mit einem dreidimensionalen Vektor l beschreiben ¹⁵⁸:

$$ax + by + c = 0 \text{ mit } l = [a \quad b \quad c]^T$$

Formel 33

Für zwei Punkte in Vektorform auf dieser Ebene gilt:

$$p_1^T l = 0 \text{ und } p_2^T l = 0 \text{ mit } p_i = [x_i \quad y_i \quad 1]^T$$

Formel 34

Über das Vektorprodukt der zwei Punkte ist eine Linie eindeutig definiert ¹⁵⁹:

$$l = p_1 \times p_2$$

Formel 35

Mit Hilfe des Kameramodells lässt sich jeder dreidimensionale Szenenpunkt anhand des zugehörigen Bildpunkts in homogenen Koordinaten und einen Skalierungsfaktor λ beschreiben ¹⁶⁰:

$$\lambda_1 \tilde{m}'_1 = M_{w1} \text{ und } \lambda_2 \tilde{m}'_2 = M_{w2} \text{ mit } \lambda_i \in (0, \infty)$$

Formel 36

Befindet sich der Ursprung des Weltkoordinatensystems im Ursprung des Koordinatensystems von Kamera 1, stehen die beiden Kameras über die folgenden euklidischen Transformationen miteinander in Beziehung ¹⁶¹:

$$M_{w2} = R M_{w1} + t = \lambda_1 R \tilde{m}'_1 + t$$

Formel 37

¹⁵⁷ Vgl. Schreer 2005, S. 75.

¹⁵⁸ Vgl. Zhang 1998.

¹⁵⁹ Vgl. Burg; Haf; Wille 2008, S. 50.

¹⁶⁰ Vgl. Schreer 2005, S. 72.

¹⁶¹ Vgl. Schreer 2005, S. 72

$$\lambda_2 \tilde{\mathbf{m}}'_2 = \lambda_1 \mathbf{R} \tilde{\mathbf{m}}'_1 + \mathbf{t}$$

Formel 38

Damit lässt sich die Epipolarlinie in Kamera 2 mit Hilfe zweier bestimmter Punkte berechnen. Beim ersten Punkt handelt es sich um den Epipol in Ansicht 2, der – wie zuvor dargelegt – die Projektion des optischen Zentrums von Kamera 1 auf der Bildebene I_2 darstellt ¹⁶². Er entspricht einer Skalierung von $\lambda_i=0$ und aus Formel 38 ergibt sich der Epipol, welcher bis auf einen Faktor dem Vektor \mathbf{t} entspricht.

$$\tilde{\mathbf{e}}'_2 = \mathbf{t}$$

Formel 39

Bei dem zweiten Punkt handelt es sich um die Abbildung eines Punktes in Ansicht 2 in unendlicher Entfernung, was einer Skalierung von $\tilde{\mathbf{m}}'_i$ mit $\lambda_i=\infty$ entspricht ¹⁶³. Dadurch kann der Vektor \mathbf{t} vernachlässigt werden und es ergibt sich die Projektion des Punktes $\tilde{\mathbf{m}}'_{2\infty}$:

$$\tilde{\mathbf{m}}'_{2\infty} = \mathbf{R} \tilde{\mathbf{m}}'_1$$

Formel 40

Mit Hilfe dieser beiden Abbildungen lässt sich die Epipolarlinie in Ansicht 2 berechnen ¹⁶⁴:

$$\mathbf{l}'_2 = \tilde{\mathbf{e}}'_2 \times \tilde{\mathbf{m}}'_{2\infty} = \mathbf{t} \times \mathbf{R} \tilde{\mathbf{m}}'_1 = \mathbf{E} \tilde{\mathbf{m}}'_1$$

Formel 41

Die Gleichung für die korrespondierende Epipolarlinie in Ansicht 1 lautet ¹⁶⁵:

$$\mathbf{l}'_1 = \mathbf{E}^T \tilde{\mathbf{m}}'_2$$

Formel 42

Ist der jeweilige Verlauf zweier korrespondierender Epipolarlinien bekannt, können daraus die zugehörigen Epipole bestimmt werden.

2. Epipol $\tilde{\mathbf{e}}'_i$: Aus der Epipolargleichung ergibt sich durch Einsatz der Formeln 41 und 42, dass für jeden Punkt in einer Ansicht der korrespondierende Punkt auf der

¹⁶² Vgl. Schreer 2005, S. 72.

¹⁶³ Vgl. Schreer 2005, S. 72.

¹⁶⁴ Vgl. Schreer 2005, S. 72.

¹⁶⁵ Vgl. Schreer 2005, S. 72.

zugehörigen Epipolarlinie der anderen Ansicht liegt. Daraus wiederum resultieren die folgenden Gleichungen ¹⁶⁶:

$$\tilde{\mathbf{m}}_1^T \mathbf{l}'_1 = 0 \text{ mit } \mathbf{l}'_1 = \mathbf{E}^T \tilde{\mathbf{m}}'_2 \text{ bzw. } \tilde{\mathbf{m}}_2^T \mathbf{l}'_2 = 0 \text{ mit } \mathbf{l}'_2 = \mathbf{E}^T \tilde{\mathbf{m}}'_1$$

Formel 43

Dadurch, dass die Epipolargleichung auch für den Epipol gilt und alle Epipolarlinien durch diesen verlaufen, gilt für alle Punkte in Ansicht 2:

$$\mathbf{E} \tilde{\mathbf{e}}'_1 = [\mathbf{t}]_{\times} \mathbf{R} \tilde{\mathbf{e}}'_1 = 0$$

Formel 44

Aus der Essential-Matrix kann damit der Epipol in Ansicht 1 aus dem rechtsseitigen eindimensionalen Nullraum berechnet werden. Dies entspricht einer Lösung des Gleichungssystems aus Formel 44 mit drei Gleichungen für drei Unbekannte mittels Gauß-Eliminierung ^{167, 168}.

Nach Division durch die dritte Komponente ergibt sich der Epipol in homogenen Koordinaten. Dementsprechend lautet die Gleichung für den Epipol in Ansicht 2:

$$\mathbf{E}^T \tilde{\mathbf{e}}'_2 = 0$$

Formel 45

Diese Gleichung ist erfüllt, wenn die Vektoren \mathbf{t} und $\mathbf{R} \tilde{\mathbf{e}}'_1$ in die gleiche Richtung zeigen, wenn also gilt:

$$\mathbf{R} \tilde{\mathbf{e}}'_1 = \mathbf{t}$$

Formel 46

Für den Epipol in Ansicht 1 ergibt sich daraus ¹⁶⁹:

$$\tilde{\mathbf{e}}'_1 = \mathbf{R}^T \mathbf{t}$$

Formel 47

3.3.2.4 Mathematische Beziehung in Bildkoordinaten

Zu Beginn dieses Kapitels wurde erläutert, dass der Zusammenhang zwischen den normierten Koordinaten der Sensorfläche und den Pixelkoordinaten des Bildes über die intrinsische Matrix \mathbf{A} hergestellt wird:

¹⁶⁶ Vgl. Schreer 2005, S. 73.

¹⁶⁷ Vgl. Schreer 2005, S. 73.

¹⁶⁸ Vgl. Burg; Haf; Wille 2008, S. 87.

¹⁶⁹ Vgl. Schreer 2005, S. 73.

$$\tilde{\mathbf{m}}_1 = \mathbf{A}_1 \tilde{\mathbf{m}}'_1 \text{ bzw. } \tilde{\mathbf{m}}_2 = \mathbf{A}_2 \tilde{\mathbf{m}}'_2$$

Entspr. Formel 18

Wird diese Beziehung mit der Epipolargleichung kombiniert, ergibt sich daraus folgender Zusammenhang ¹⁷⁰:

$$\tilde{\mathbf{m}}_2^T \mathbf{A}_2^{-T} \mathbf{E} \mathbf{A}_1^{-1} \tilde{\mathbf{m}}_1 = \tilde{\mathbf{m}}_2^T \mathbf{F} \tilde{\mathbf{m}}_1 = 0 \text{ mit } \mathbf{F} = \mathbf{A}_2^{-T} \mathbf{E} \mathbf{A}_1^{-1}$$

Formel 48

\mathbf{F} repräsentiert eine singuläre 3×3-Matrix und wird als Fundamental-Matrix bezeichnet. Sie enthält die intrinsischen Parameter beider Kameras sowie die extrinsischen Parameter der euklidischen Transformation ¹⁷¹. Dadurch beschreibt \mathbf{F} die Epipolar-geometrie vollständig in Pixelkoordinaten. Die Gleichungen für die beiden Epipolarlinien lauten unter Nutzung der Fundamental-Matrix ¹⁷²:

$$\mathbf{l}_2 = \mathbf{F} \tilde{\mathbf{m}}_1 \text{ und } \mathbf{l}_1 = \mathbf{F}^T \tilde{\mathbf{m}}_2$$

Formel 49

Die Gleichungen für die beiden Epipole in Bildkoordinaten können unter Nutzung der intrinsischen Matrizen umformuliert werden:

$$\tilde{\mathbf{e}}_2 = \mathbf{A}_2 \mathbf{t} \text{ und } \tilde{\mathbf{e}}_1 = \mathbf{A}_1 \mathbf{R}^T \mathbf{t}$$

Formel 50

Darüber hinaus kann die Definition der Fundamental-Matrix als das Produkt aus einer nicht-singulären (regulären) Matrix und einem Vektor in Form einer antisymmetrischen Matrix notiert werden ¹⁷³:

$$\mathbf{F} = \mathbf{A}_2^{-T} [\mathbf{t}]_{\times} \mathbf{R} \mathbf{A}_1^{-1} = [\mathbf{A}_2 \mathbf{t}]_{\times} \mathbf{A}_2 \mathbf{R} \mathbf{A}_1^{-1} \text{ mit } [\mathbf{A}_2 \mathbf{t}]_{\times} = \mathbf{A}_2^{-T} [\mathbf{t}]_{\times}$$

Formel 51

Daraus ergibt sich unter Verwendung der Definition für den Epipol in Ansicht 2:

$$\mathbf{F} = [\tilde{\mathbf{e}}_2]_{\times} \mathbf{A}_2 \mathbf{R} \mathbf{A}_1^{-1}$$

Formel 52

Entsprechend der Gleichung für die Epipole in Kamerakoordinaten gelten für die Fundamental-Matrix und die Epipole in Bildkoordinaten die

¹⁷⁰ Vgl. Schreer 2005, S. 74.

¹⁷¹ Vgl. Schreer 2005, S. 74.

¹⁷² Vgl. Faugeras; Loung 2001, S. 261.

¹⁷³ Vgl. Faugeras; Loung 2001, S. 285.

folgenden Beziehungen ¹⁷⁴:

$$\mathbf{F} \tilde{\mathbf{e}}_1 = 0 \text{ und } \mathbf{F}^T \tilde{\mathbf{e}}_2 = 0$$

Formel 53

Wird die Epipolargeometrie durch die Fundamental-Matrix ausgedrückt, so enthält sie die Epipole in den Ansichten 1 und 2. Sie können aus \mathbf{F} durch Lösen des rechtsseitigen Nullraumes bestimmt werden ¹⁷⁵. Wie auch die Essential-Matrix beschreibt die Fundamental-Matrix vollständig die geometrische Beziehung zwischen korrespondierenden Punkten in den zwei Ansichten eines Stereokamerasystems. Im Gegensatz zu \mathbf{E} drückt \mathbf{F} die Beziehung in Bildkoordinaten aus und nicht in Sensorkoordinaten, weshalb hier von unkalibrierten Kameras gesprochen wird. Beide Matrizen weisen Bedingungen auf, die für weitere Berechnungen wichtig sind.

1. Bedingung für \mathbf{E} : Die Determinante einer antisymmetrischen Matrix mit ungerader Dimension ist Null, daher gilt dies auch für die Determinante der Essential-Matrix ¹⁷⁶:

$$\det(\mathbf{E}) = \det([\mathbf{t}]_{\times}) \det(\mathbf{R}) = 0$$

Formel 54

2. Bedingung für \mathbf{E} : Aufgrund von $\det(\mathbf{E})=0$ existieren nur zwei linear unabhängige Zeilen- oder Spaltenvektoren. Somit hat die Matrix den Rang $Rg(\mathbf{E})=2$. Die beiden von Null verschiedenen Eigenwerte der Essential-Matrix haben den gleichen Wert ¹⁷⁷.

1. Bedingung für \mathbf{F} : Da die Determinante der Essential-Matrix verschwindet, gilt für \mathbf{F} ebenfalls $\det(\mathbf{F})=0$.

2. Bedingung für \mathbf{F} : Somit besitzt auch die Fundamental-Matrix den Rang $Rg(\mathbf{F})=2$ ¹⁷⁸. Durch die Eigenschaft $\det(\mathbf{F})=0$ und weil sie nur bis auf einen Skalierungsfaktor definiert ist, sind von den neun Elementen der 3×3 -Matrix nur sieben Elemente voneinander unabhängig ¹⁷⁹.

¹⁷⁴ Vgl. Faugeras; Loung 2001, S. 263.

¹⁷⁵ Vgl. Schreer 2005, S. 75.

¹⁷⁶ Vgl. Wikipedia: Schiefsymmetrische Matrix.

¹⁷⁷ Vgl. Faugeras; Loung 2001, S. 284.

¹⁷⁸ Vgl. Zhang 1998.

¹⁷⁹ Vgl. Hartley; Zisserman 2003, S. 246.

3.3.3 Achsparalleles System

3.3.3.1 Modellierung

Auch hier existieren zwei optische Zentren bzw. Brennpunkte C_1 und C_2 sowie zwei Bildebenen I_1 und I_2 . Die Brennpunkte und die Bildebenen weisen hinsichtlich ihrer Position lediglich jeweils einen horizontalen Versatz auf. Die Bildpunkte m_1 und m_2 liegen auf gleicher Höhe, d.h. der jeweilige Zeilenindex n hat im linken und rechten Bild den gleichen Wert. Der Wert p stellt die Länge einer Strecke dar, die durch M_w verläuft und die Linie der Stereobasis SB im rechten Winkel schneidet. Die unterschiedlichen Perspektiven der beiden Kameras führen zu einem horizontalen Versatz zwischen m_1 und m_2 , der Disparität δ ¹⁸⁰.

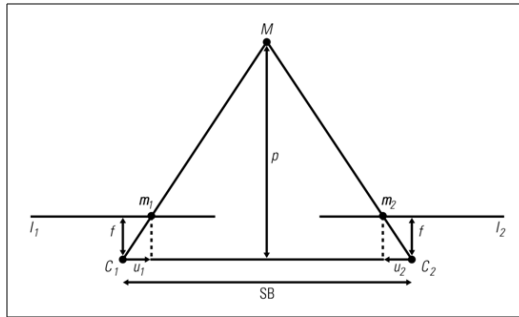


Abbildung 15: Achsparalleles System

Die Disparität wird in Pixel angegeben, da sie aus den Bildkoordinaten berechnet wird¹⁸¹:

$$\delta = u_1 - u_2$$

Formel 55

Entsprechend der Abbildung lässt sich mit Hilfe des Strahlensatzes die Aussage treffen, dass das Verhältnis zwischen p und f gleich dem Verhältnis zwischen SB und $\delta * d_u$ (in mm) sein muss. Der Parameter d_u repräsentiert einen Skalierungsfaktor, welcher die Pixelwerte der Disparität in metrische Werte umrechnet. Die entsprechende Gleichung lautet dann:

$$\frac{p}{f} = \frac{SB}{\delta * d_u}$$

Formel 56

¹⁸⁰ Vgl. Schmidt 2009, S. 498.

¹⁸¹ Vgl. Schreer 2005, S. 67.

Bei einer Auflösung dieser Gleichung nach p ergibt sich:

$$p = \frac{SB * f}{\delta * d_u}$$

Formel 57

Sind die einzelnen Variablen bekannt, lässt sich daraus die Raumtiefe p des Punktes M_w bestimmen. Es wird von einem System ausgegangen, bei dem die Stereobasis und die Brennweite bekannt sind, daher liegt die Herausforderung in der Bestimmung des Disparitätswerts. Raumtiefe und Disparität stehen bei achsparallelen Systemen in einem reziproken Verhältnis, für sehr nahe Punkte wird die Disparität größer, bei weit entfernten Punkten geht die Disparität gegen Null ¹⁸².

3.3.3.2 Mathematische Beziehungen

Die euklidische Transformation vereinfacht sich in einem achsparallelen System zu $R=I$ ¹⁸³. Unter der Voraussetzung, dass die intrinsischen Matrizen beider Kameras identisch sind ($A_1=A_2$) ergibt sich für die Fundamental-Matrix aus Formel 51:

$$F = [\tilde{e}_2]_{\times} A_2 A_1^{-1} = [\tilde{e}_2]_{\times}$$

Formel 58

Für den Epipol in Ansicht 2 ergeben sich aus Formel 50 die Koordinaten:

$$e_2 = (1,0,0)^T$$

Formel 59

Aufgrund der Tatsache, dass die dritte Koordinate Null ist, handelt es sich bei dem Epipol um einen Punkt im Unendlichen ¹⁸⁴. Da ein Epipol die (theoretische) Abbildung eines optischen Zentrums in der jeweils anderen Projektionsebene ist, ist diese Gegebenheit bei einem Achsparallelen System leicht nachzuvollziehen. Die Fundamentalmatrix besitzt dann folgende Form:

$$F = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$$

Formel 60

¹⁸² Vgl. Jähne 2005, S. 232.

¹⁸³ Vgl. Hartley; Zisserman 2003, S. 249.

¹⁸⁴ Vgl. Hartley; Zisserman 2003, S. 242.

Diese zwei Freiheitsgrade korrespondieren mit der Lage der Epipole. Damit ergibt sich für einen beliebigen Bildpunkt in Ansicht 1 eine Epipolarlinie in Ansicht 2, welche nur noch die vertikale Koordinate des Bildpunktes enthält ¹⁸⁵:

$$l_2 = F\tilde{m}_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ v \end{bmatrix}$$

Formel 61

Über die bereits bekannte Geradendefinition

$$au + bv + c = 0 \text{ mit } m = \frac{-a}{b} \text{ und } d = \frac{-c}{b}$$

Entspr. 33

ergibt sich für diese Epipolarlinie die Steigung $m=0$ und der Achsenabschnitt $d=v$. Dies ist die horizontale Epipolarlinie mit der gleichen v -Koordinate des Bildpunktes in der anderen Projektionsebene ¹⁸⁶.

Über die jeweiligen intrinsischen und extrinsischen Parameter lässt sich die Epipolargeometrie zweier Kameras bestimmen. Je nach Anwendung kann diese dann über die Essential- oder Fundamental-Matrix ausgedrückt werden. Über die Epipolargleichung stehen die Bildpunkte in Pixelkoordinaten in direktem Zusammenhang mit der Fundamental-Matrix. In der Regel sind bei der Aufnahme die intrinsischen und extrinsischen Kameraparameter nicht bekannt und müssen bestimmt werden. Dies geschieht über Verfahren, die unter der Bezeichnung Kalibrierung zusammengefasst werden. Hierzu werden hauptsächlich sog. Kalibrierungskörper verwendet. Diese Körper weisen Muster auf, für die die exakten 3D-Koordinaten vorliegen ¹⁸⁷. Durch geeignete Bildverarbeitungsverfahren können die Muster identifiziert und die Positionen ihrer Abbildungen in Bildkoordinaten bestimmt werden. Mit diesen Informationen lassen sich die Kameraparameter zuverlässig schätzen. Die unterschiedlichen Kalibrierungsverfahren sind unumgänglich zur Bestimmung der Parameter einer einzelnen Kamera. Die manuelle Kalibrierung soll in dieser Arbeit nicht näher behandelt werden, da die Fundamental-Matrix bei einem Stereosystem auch ohne Kenntnis der expliziten Kameraparameter direkt aus den Stereoansichten gewonnen werden kann ¹⁸⁸. Da nur Informationen aus den projizierten Abbildungen der Szene vorliegen, wird diese Konstellation als projektive Stereogeometrie und die Fundamental-Matrix als eine projektive Größe bezeichnet.

¹⁸⁵ Vgl. Faugeras; Loung 2001, S. 289.

¹⁸⁶ Vgl. Schreer 2005, S. 77.

¹⁸⁷ Vgl. Hartley; Zisserman 2003, S. 182.

¹⁸⁸ Vgl. Schreer 2005, S. 21f.

3.3.4 Gewinnung der Fundamental-Matrix

Nachfolgend findet sich noch einmal die Gleichung für korrespondierende Punkte in den beiden Ansichten unter Verwendung der Fundamental-Matrix:

$$\tilde{\mathbf{m}}_2^T \mathbf{F} \tilde{\mathbf{m}}_1 = 0 \text{ mit } \mathbf{F} = \mathbf{A}_2^{-T} \mathbf{E} \mathbf{A}_1^{-1}$$

Entspr. Formel 48

Die Schätzung der Fundamental-Matrix kann über lineare oder nichtlineare Optimierungsverfahren erreicht werden.

3.3.4.1 Lineare Verfahren

Bei linearen Schätzungsverfahren hat der Wertebereich der Messdaten Einfluss auf die Robustheit der Schätzung. Die Messdaten sind in diesem Fall repräsentiert durch die Bildkoordinaten, welche allerdings sowohl in horizontaler als auch vertikaler Richtung einen Wertebereich von mehreren hundert Pixeln aufweisen. Um einen sinnvollen Wertebereich zu definieren, wird daher vor der Schätzung eine sog. isotrope Normierung der Messdaten vorgenommen¹⁸⁹. Isotrop bedeutet in diesem Fall eine von der Richtung unabhängige Skalierung. Die Normierung wird als Vorverarbeitungsschritt bei linearen Verfahren explizit empfohlen, da der entsprechende Algorithmus relativ einfach zu implementieren ist und das Ergebnis der Schätzung positiv beeinflusst¹⁹⁰.

Die Epipolargleichung kann als lineare homogene Gleichung mit den neun Unbekannten der Koeffizienten f_{ij} der Fundamental-Matrix umformuliert werden¹⁹¹.

$$\mathbf{u}_i^T \mathbf{f} = 0 \text{ mit}$$

$$\mathbf{u}_i = [u_{i2} \ u_{i1}, u_{i2} \ v_{i1}, u_{i2} \ v_{i2} \ u_{i1}, v_{i2} \ v_{i1}, v_{i2} \ u_{i1}, v_{i1}, 1]^T \text{ und}$$

$$\mathbf{f} = [f_{11}, f_{12}, f_{13}, f_{21}, f_{22}, f_{23}, f_{31}, f_{32}, f_{33}]^T$$

Formel 62

Für n gegebene Punktkorrespondenzen ergibt sich ein lineares Gleichungssystem:

$$\mathbf{U}_n \mathbf{f} = 0 \text{ mit } \mathbf{U}_n = \begin{bmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_n^T \end{bmatrix}$$

Formel 63

¹⁸⁹ Vgl. Schreer 2005, S. 21f.

¹⁹⁰ Vgl. Hartley; Zisserman 2003, S. 281f.

¹⁹¹ Vgl. Faugeras; Loug 2001, S 318.

Aufgrund der Tatsache, dass die Fundamental-Matrix bis auf einen Skalierungsfaktor definiert ist, kann mit dem Gleichungssystem bei genau acht gegebenen Punktkorrespondenzen eine eindeutige Lösung für \mathbf{F} berechnet werden¹⁹². In diesem Fall muss zur Lösung des Gleichungssystems eine Singularwertzerlegung durchgeführt werden. Daraus ergibt sich eine Pseudoinverse $\mathbf{U}_n^{(-)}$ der Matrix \mathbf{U}_n , mit der das Gleichungssystem gelöst und die Elemente der Fundamentalmatrix bestimmt werden können. Dieses Konzept wird als 8-Punkt-Algorithmus bezeichnet¹⁹³.

Oft weisen die Korrespondenzen jedoch nicht die erforderliche Genauigkeit auf, daher muss die Matrix so genau wie möglich geschätzt werden. Um die Ungenauigkeiten zu verringern, können mehr als acht Korrespondenzen herangezogen werden. Unter Nutzung der Epipolargleichung kann ein quadratischer Ansatz aufgestellt werden¹⁹⁴:

$$\min_f \sum_i \|\mathbf{U}_n \mathbf{f}\|^2$$

Formel 64

Der Vektor \mathbf{f} ist bis auf einen Skalierungsfaktor definiert. Um die Lösung $\mathbf{f}=\mathbf{0}$ auszuschließen, können zusätzliche erforderliche Bedingungen definiert werden. Beim ersten Ansatz wird einer der Koeffizienten der Fundamental-Matrix auf Eins gesetzt und das Gleichungssystem über die Methode des kleinsten quadratischen Fehlers für ein reduziertes Gleichungssystem mit acht Gleichungen gelöst¹⁹⁵. Ein zweiter Ansatz besteht in der Festlegung der Norm des Vektors \mathbf{f} ¹⁹⁶. Im Gegensatz zum ersten Ansatz werden hier alle Koeffizienten von \mathbf{F} gleich behandelt.

Eine Lösung, die durch den 8-Punkt-Algorithmus ermittelt wird, vernachlässigt i. A. die Bedingung, dass die Fundamental-Matrix singular ist und den Rang $Rg(\mathbf{F})=2$ besitzt. Die Singularitätsbedingung kann erzwungen werden, indem für \mathbf{F} eine Singularwertzerlegung durchgeführt wird und die Matrix dann durch eine andere Matrix \mathbf{F}' ersetzt wird¹⁹⁷. \mathbf{F}' erfüllt die Rang-Bedingung und kommt der geschätzten Matrix \mathbf{F} am nächsten.

Die Fundamental-Matrix besitzt sieben unabhängige Parameter. Das Gleichungssystem $\mathbf{U}_n \mathbf{f}=\mathbf{0}$ mit sieben Punktkorrespondenzen liefert eine 7×9 -Matrix \mathbf{U} mit dem Rang $Rg(\mathbf{U})=7$ ¹⁹⁸. Die Lösung des Gleichungssystems ist der zweidimensionale Nullraum:

192 Vgl. Schreer 2005, S. 80.

193 Vgl. Sur; Noury; Berger 2008.

194 Vgl. Zhang 1998.

195 Vgl. Schreer 2005, S. 81.

196 Vgl. Zhang 1998.

197 Vgl. Hartley; Zisserman 2003, S. 282.

198 Vgl. Hartley; Zisserman 2003, S. 281.

$$\mathbf{F}' = \alpha \mathbf{F}_1 + (1 - \alpha) \mathbf{F}_2$$

Formel 65

Wird die Rang-Bedingung eingehalten, ergibt sich eine Gleichung dritten Grades bezüglich des skalaren Parameters α :

$$\det(\mathbf{F}') = \det(\alpha \mathbf{F}_1 + (1 - \alpha) \mathbf{F}_2) = 0$$

Formel 66

Die Gleichung liefert genau eine oder drei reelle Lösungen für α ¹⁹⁹. Die korrekte Lösung kann unter Nutzung von Formel 65 ermittelt werden. Dieses Verfahren wird 7-Punkt-Algorithmus genannt und liefert eine Fundamental-Matrix bei der Minimalanzahl von sieben Punktkorrespondenzen.

3.3.4.2 Nichtlineare Verfahren

Lineare Verfahren liefern akzeptable Ergebnisse, wenn die Punktkorrespondenzen hinreichend exakt sind. Nach Schreer liefern jedoch nichtlineare Verfahren aus der numerischen Optimierung bessere Resultate²⁰⁰. Es wird dabei ein Parametervektor definiert, der die gesuchten Größen, d.h. die Elemente der Fundamental-Matrix, enthält. Mit N als Parameteranzahl wird damit eine Suche im N -dimensionalen Suchraum durchgeführt. Bei nichtlinearen Verfahren wird ein geeigneter Startwert benötigt, aus diesem Grund wird i.d.R. zunächst über ein lineares Verfahren eine Fundamental-Matrix geschätzt und anschließend über ein nichtlineares Verfahren optimiert²⁰¹. Als Beispiel ist hier der sog. M-Schätzer zu nennen. Dieses Verfahren basiert auf der Idee, eine Funktion durch eine andere Funktion zu ersetzen, bei der Ausreißer oder Abweichungen weniger stark berücksichtigt werden. In diesem konkreten Fall wird die Methode des kleinsten quadratischen Fehlers durch eine Gütefunktion ersetzt, die eine iterative Gewichtung der Messwerte beinhaltet²⁰². Dadurch kann der Einfluss von Ungenauigkeiten bei den Punktkorrespondenzen reduziert werden.

Werden Punktkorrespondenzen automatisch berechnet, resultiert aus der fehlenden Kontrolle durch einen Anwender eine hohe Anfälligkeit für Ausreißer, d.h. für ungenau lokalisierte Punktkorrespondenzen und falsch zugeordnete Korrespondenzen. Die Epipolarbedingung kann generell als Ähnlichkeitskriterium zur Bestimmung von Punktkorrespondenzen genutzt werden. Das Problem in diesem Fall ist, dass diese zum notwendigen Zeitpunkt noch nicht verfügbar ist, da sie über die Fundamental-Matrix beschreiben werden muss, welche wiederum das angestrebte Ergebnis der

¹⁹⁹ Vgl. Eschenburg 2006, S. 31.

²⁰⁰ Vgl. Schreer 2005, S. 85.

²⁰¹ Vgl. Schreer 2005, S. 85.

²⁰² Vgl. Faugeras; Loug 2001, S. 330.

Schätzung ist ²⁰³. Einige Verfahren zur Schätzung der projektiven Stereogeometrie bieten daher nicht nur eine optimale Anpassung der Fundamental-Matrix an die Messdaten, sondern eliminieren gleichzeitig auch Abweichungen innerhalb der Messdaten, d.h. den Punktkorrespondenzen.

Der sog. Random Sample Consensus-Algorithmus (kurz: RANSAC) stellt ein solches Verfahren dar ²⁰⁴. Im Gegensatz zu klassischen Schätzverfahren werden dabei nicht alle verfügbaren Messwerte berücksichtigt, sondern es werden nach dem Zufallsprinzip aus der Menge der Messungen so viele Werte entnommen, wie für die Berechnung der Parameter des Modells notwendig sind. Entsprechend den Erläuterungen im vorangegangenen Abschnitt werden für die Fundamental-Matrix sieben Werte benötigt. Mittels Gütekriterium erfolgt eine Bewertung des geschätzten Modells und anhand des Resultats wird entschieden, ob die Messungen verworfen oder weiterverwendet werden. Auf diese Weise erhält man zwar weniger Messwerte, dafür ist jedoch gewährleistet, dass keine Ausreißer enthalten sind ²⁰⁵. Um das Modell an die verbleibenden Werte anzupassen, wird zusätzlich eine nichtlineare Optimierung des Parametervektors vorgenommen, ähnlich der Vorgehensweise, die beim o.g. M-Schätzer verfolgt wird. Problematisch beim RANSAC-Algorithmus ist die Definition eines geeigneten Schwellenwerts, welcher bestimmt, ab wann die Schätzung der Fundamental-Matrix exakt genug ist ²⁰⁶. Eine Lösung hierzu besteht in einem Algorithmus, der den Schwellenwert bei jedem Iterationsschritt anpasst.

3.3.5 Beziehungen zwischen zwei Stereoansichten

Mit der Epipolargleichung wird eine Beziehung hergestellt zwischen korrespondierenden Punkten in zwei unterschiedlichen Ansichten. Nun muss untersucht werden, wie die Transformation von Abbildungen in einer Ansicht in die zugehörige Abbildung der anderen Ansicht verläuft. Wie bereits erläutert, führen Objektpunkte unabhängig von ihrer Entfernung zur Projektionsebene in einer Ansicht zu ein und derselben Abbildung, solange sie sich auf der Geraden befinden, welche durch den Bildpunkt und den Brennpunkt verläuft. In der anderen Ansicht jedoch ergeben sich daraus unterschiedliche Abbildungen, daher hängt diese Beziehung von der Raumtiefe des Objektpunktes ab. Mit Hilfe der sog. allgemeinen Disparitätsgleichung lässt sich ein eindeutiger Zusammenhang zwischen den Abbildungen eines Objektpunktes in zwei Ansichten beschreiben ²⁰⁷:

$$\tilde{\mathbf{m}}_2 = \mathbf{A}_2 \mathbf{R} \mathbf{A}_1^{-1} \tilde{\mathbf{m}}_1 + \frac{1}{Z_1} \mathbf{A}_2 \mathbf{t}$$

Formel 67

²⁰³ Vgl. Schreer 2005, S. 86.

²⁰⁴ Vgl. Fischler; Bolles 1981

²⁰⁵ Vgl. Fischler; Bolles 1981.

²⁰⁶ Vgl. Hartley; Zisserman 2003, S. 118.

²⁰⁷ Vgl. Schreer 2005, S. 94.

Diese Gleichung ist nichtlinear. Für die Zusammenhänge bei der Projektion eines Objektpunktes in zwei Ansichten ist die Frage von Bedeutung, unter welchen Bedingungen die Gleichung eine lineare Transformation darstellt. Die erste Möglichkeit besteht darin, dass der Translationsvektor \mathbf{t} gleich Null wird. In diesem Fall ist die Position der optischen Zentren beider Ansichten identisch und die Projektionsebenen sind lediglich gedreht²⁰⁸. Abgesehen von der Unmöglichkeit einer praktischen Umsetzung liefert dieser Fall keine zweite Perspektive auf die Szene, ist also für eine stereoskopische Aufnahme ungeeignet. Die zweite mögliche Bedingung besteht darin, dass die Objektpunkte sehr weit von den Kameras entfernt sind, d.h. der Term $1/Z_1$ geht für $Z_1 \rightarrow \infty$ gegen Null. Die Gleichung stellt dann eine lineare Transformation dar²⁰⁹:

$$\tilde{\mathbf{m}}_2 = \mathbf{A}_2 \mathbf{R} \mathbf{A}_1^{-1} \tilde{\mathbf{m}}_1 = \mathbf{H}_\infty \tilde{\mathbf{m}}_1 \text{ mit } \mathbf{H}_\infty = \mathbf{A}_2 \mathbf{R} \mathbf{A}_1^{-1}$$

Formel 68

Die lineare Transformationsmatrix wird als \mathbf{H}_∞ bezeichnet, da sie nur für Punkte im Unendlichen gültig ist. Dies bedeutet, dass Positionsänderungen sehr weit entfernter Objekte keine Veränderung in der Abbildung mit sich bringen²¹⁰.

Werden koplanare Punkte (d.h. Punkte, die alle in der gleichen Ebene liegen) auf eine andere Ebene projiziert, kann die Projektion über eine lineare Transformation beschrieben werden. Dies wird als Homographie bezeichnet, die zugehörige Homographie-Matrix \mathbf{H} ist nicht-singulär, d.h. invertierbar²¹¹. \mathbf{H} ist eine 3×3 -Matrix und bis auf einen Skalierungsfaktor definiert, daraus ergeben sich 8 Freiheitsgrade. Die Kombination zweier projektiver Transformationen liefert eine neue projektive Transformation. Die Projektion von Objektpunkten einer Ebene π in die Bildebene von Kamera 1 stellt ebenso wie die Projektion der Objektpunkte in die Bildebene von Kamera 2 eine projektive Transformation dar. Aufgrund der Homographie stellt dann die Transformation der Bildpunkte einer Ansicht in die andere Ansicht ebenfalls eine lineare Transformation dar.

208 Vgl. Faugeras; Loun 2001, S. 286.

209 Vgl. Faugeras; Loun 2001, S. 281.

210 Vgl. Schreer 2005, S. 95.

211 Vgl. Schreer 2005, S. 17.

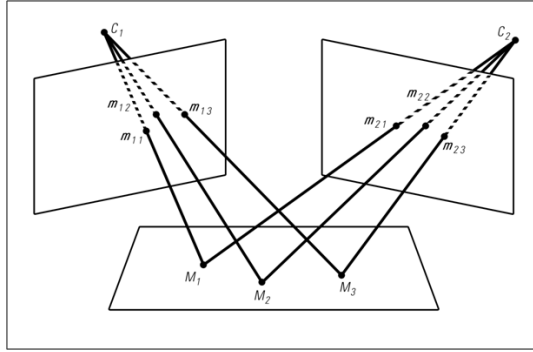


Abbildung 16: Homographie

Die Ebene π kann über ihren Normalenvektor \mathbf{n} und den Abstand d zum optischen Zentrum von Kamera 1 C_1 definiert werden ²¹²:

$$\pi = [\mathbf{n}^T - d]^T$$

Formel 69

Unter der Voraussetzung, dass π nicht durch C_1 verläuft, ergibt sich für die Homographie-Matrix ²¹³:

$$\mathbf{H} = \mathbf{H}_\infty + \tilde{\mathbf{e}}_1 \frac{\mathbf{n}^T \mathbf{A}_2^{-1}}{d} = \mathbf{A}_1 \mathbf{R} \mathbf{A}_2^{-1} + \mathbf{A}_1 \mathbf{t} \frac{\mathbf{n}^T \mathbf{A}_2^{-1}}{d}$$

Formel 70

Lautet die Voraussetzung, dass π nicht durch C_2 verläuft, kann die inverse Homographie-Matrix genutzt werden:

$$\mathbf{H}^{-1} = \mathbf{H}_\infty^{-1} + \tilde{\mathbf{e}}_2 \frac{\mathbf{n}^T \mathbf{A}_1^{-1}}{d}$$

Formel 71

Die dritte Möglichkeit besteht darin, dass π beide Brennpunkte einbezieht und somit die Epipolarebene darstellt ($\pi = \Pi$). In diesem Fall ist die Homographie nicht definiert,

²¹² Vgl. Schreer 2005, S. 96.

²¹³ Vgl. Schreer 2005, S. 97.

da die Epipolarebene die optischen Strahlen des Objektpunktes in beiden Bildebenen enthält. Dadurch besteht die Möglichkeit, dass mehrere Abbildungen in einer Ansicht mit einer einzigen Abbildung in der anderen Ansicht korrespondieren, wodurch die Eindeutigkeit der Zuordnung nicht mehr gewährleistet ist. Kann eine Homographie-Matrix \mathbf{H} aufgestellt werden, so existiert für koplanare Objektpunkte einer Ebene eine lineare Transformation, welche jeden Punkt \mathbf{m}_1 in der Bildebene I_1 in den korrespondierenden Bildpunkt \mathbf{m}_2 in der Bildebene I_2 abbildet ²¹⁴.

3.3.5.1 Homographie und Fundamental-Matrix

Wie zuvor bereits erläutert, sind die Epipole unabhängig von der Position der Objektpunkte. Aus diesem Grund können beide Epipole über eine beliebige Homographie in Beziehung gesetzt werden ²¹⁵:

$$\tilde{\mathbf{e}}_2 = \mathbf{H} \tilde{\mathbf{e}}_1$$

Formel 72

Die Epipolarlinien können ebenfalls mit einer beliebigen Homographie ineinander überführt werden:

$$\tilde{\mathbf{l}}_2 = \mathbf{H}^{-T} \tilde{\mathbf{l}}_1$$

Formel 73

Zur Beschreibung des Zusammenhangs zwischen Homographie und Fundamental-Matrix wird die bereits erwähnte Epipolarbedingung herangezogen. Kombiniert man diese mit der Homographie-Beziehung für korrespondierende Abbildungen $\tilde{\mathbf{m}}_2 = \mathbf{H} \tilde{\mathbf{m}}_1$, lautet die Epipolarbedingung:

$$(\mathbf{H} \tilde{\mathbf{m}}_1)^T \mathbf{F} \tilde{\mathbf{m}}_1 = \tilde{\mathbf{m}}_1^T \mathbf{H}^T \mathbf{F} \tilde{\mathbf{m}}_1 = 0$$

Formel 74

Damit diese Gleichung erfüllt ist, muss $\mathbf{H}^T \mathbf{F}$ antisymmetrisch sein. Durch Berücksichtigung dieser Eigenschaft ergeben sich weitere Beziehungen ^{216, 217}:

$$(\mathbf{H}^T \mathbf{F})^T = -\mathbf{H}^T \mathbf{F} \Rightarrow \mathbf{F}^T \mathbf{H} + \mathbf{H}^T \mathbf{F} = 0$$

Formel 75

Um den Bezug herzustellen zum Zusammenhang zwischen Homographie und Funda-

²¹⁴ Vgl. Schreer 2005, S. 97.

²¹⁵ Vgl. Schreer 2005, S. 97.

²¹⁶ Vgl. Hartley; Zisserman 2003, S. 327.

²¹⁷ Vgl. Faugeras; Loug 2001, S. 270.

mental-Matrix, wird die Homographie-Beziehung mit der Gleichung für die Epipolarlinie in Ansicht 2 kombiniert:

$$l_2 = \tilde{e}_2 \times \tilde{m}_2 = \tilde{e}_2 \times (H \tilde{m}_1) = [\tilde{e}_2]_{\times} * H \tilde{m}_1$$

Formel 76

Aufgrund der Definition der Epipolarlinie $l_2 = F \tilde{m}_1$ ergibt sich für den Zusammenhang zwischen Homographie und Fundamental-Matrix ²¹⁸:

$$F = [\tilde{e}_2]_{\times} * H$$

Formel 77

Die Fundamental-Matrix lässt sich also über den Epipol und eine beliebige nicht-singuläre 3×3-Matrix faktorisieren. Die Fundamental-Matrix beschreibt die relative Beziehung zwischen zwei Kameras, daher kann das projektive Koordinatensystem der beiden Kameras frei gewählt werden. Der Einfachheit halber kann die Projektionsmatrix von Kamera 1 folgendermaßen definiert werden ²¹⁹:

$$P_1 = [I \mid 0]$$

Formel 78

Das Weltkoordinatensystem wird dadurch in Kamera 1 gelegt und die intrinsische Matrix ist die Einheitsmatrix, d.h. sie ist quadratisch, sämtliche Werte der Hauptdiagonalen sind Eins und die restlichen Werte sind Null. Beim Vorhandensein einer Fundamental-Matrix, die entsprechend der oben stehenden Gleichung zerlegt wurde, ergibt sich für die Projektionsmatrix der zweiten Kamera ²²⁰:

$$P_2 = [H \mid \tilde{e}_2]$$

Formel 79

3.3.5.2 Berechnung der Homographie

Durch die Beziehung zwischen der Fundamental-Matrix und der Homographie ist es möglich, eine der beiden Größen bei Kenntnis der jeweils anderen zu bestimmen. Es folgt die Berechnung der Homographie bei bekannter Fundamental-Matrix. Die

²¹⁸ Vgl. Faugeras; Loung 2001, S. 271.

²¹⁹ Vgl. Zhang 1998.

²²⁰ Vgl. Zhang 1998.

Summe zweier antisymmetrischer Matrizen ergibt mindestens eine symmetrische Matrix. Da $\mathbf{H}^T \mathbf{F}$ eine antisymmetrische Matrix ist, ergeben sich aus Formel 75 sechs homogene Gleichungen zur Berechnung der unbekannten Parameter der Homographie \mathbf{H} . Die Gleichung enthält lediglich projektive Größen, die bis auf einen Skalierungsfaktor definiert sind, daher verbleiben nur fünf linear unabhängige Gleichungen²²¹. Um die acht Parameter der Homographie zu berechnen, fehlen damit noch drei Gleichungen. Für zwei korrespondierende Punkte in Ansicht 1 und 2 ergeben sich aufgrund der Epipolargeometrie folgende Vektoren: $\tilde{\mathbf{m}}_2, \mathbf{F}\tilde{\mathbf{m}}_1, \mathbf{H}\tilde{\mathbf{m}}_1$. Das Spatprodukt daraus lautet:

$$(\tilde{\mathbf{m}}_2, \mathbf{F}\tilde{\mathbf{m}}_1, \mathbf{H}\tilde{\mathbf{m}}_1) = \tilde{\mathbf{m}}_2 * (\mathbf{F}\tilde{\mathbf{m}}_1 \times \mathbf{H}\tilde{\mathbf{m}}_1) = \tilde{\mathbf{m}}_2 * (\mathbf{F}\tilde{\mathbf{m}}_1 \times \tilde{\mathbf{m}}_2) = 0$$

Formel 80

Dieses Spatprodukt verschwindet aufgrund der Tatsache, dass zwei der drei Vektoren gleich sind. Daraus ergibt sich eine skalare Gleichung für eine Punktkorrespondenz, die die Homographie und die Fundamental-Matrix enthält. Mit drei verschiedenen nicht-kollinearen Punktkorrespondenzen ergeben sich die drei benötigten Gleichungen zur Berechnung der acht freien Parameter der Homographie²²².

$$\mathbf{F}^T \mathbf{H} + \mathbf{H}^T \mathbf{F} = 0$$

Formel 75

$$(\tilde{\mathbf{m}}_{21}, \mathbf{F}\tilde{\mathbf{m}}_{11}, \mathbf{H}\tilde{\mathbf{m}}_{11}) = 0$$

Formel 81

$$(\tilde{\mathbf{m}}_{22}, \mathbf{F}\tilde{\mathbf{m}}_{12}, \mathbf{H}\tilde{\mathbf{m}}_{12}) = 0$$

Formel 82

$$(\tilde{\mathbf{m}}_{23}, \mathbf{F}\tilde{\mathbf{m}}_{13}, \mathbf{H}\tilde{\mathbf{m}}_{13}) = 0$$

Formel 83

Bei bekannter Fundamental-Matrix werden zur Bestimmung der Homographie drei Punktkorrespondenzen benötigt²²³. Da drei Punkte im Raum eine Ebene aufspannen, muss es nach dem bisher Erläuterten eine Homographie für die korrespondierenden Punkte in den beiden Stereoansichten geben. Bei dem soeben beschriebenen Verfahren werden genau diese drei Punktkorrespondenzen benötigt, woraus sich die gesuchte Homographie ergibt.

221 Vgl. Hartley; Zisserman 2003, S. 327f.

222 Vgl. Schreer 2005, S. 100f.

223 Vgl. Faugeras; Loung 2001, S. 273.

3.3.5.3 Berechnung der Fundamental-Matrix

Im vorherigen Abschnitt wurde der Zusammenhang zwischen Homographie und Fundamental-Matrix mit $F = [e_2]_{\pi} \cdot H$ definiert. Bei bekannter Homographie lässt sich also die Fundamental-Matrix berechnen, wenn der Epipol in Ansicht 2 bekannt ist. Dies wird bewerkstelligt, indem zwei Epipolarlinien in der Ansicht ermittelt werden, da der gesuchte Epipol den Schnittpunkt der beiden Linien darstellt. Nach Schreer liegen über die Berechnung der Homographie bereits vier Punktkorrespondenzen vor, die die Ebene π aufspannen²²⁴. Diese Punkte werden logischerweise in beiden Ansichten abgebildet. Nun wird eine weitere Punktkorrespondenz (m_{15}, m_{25}) gesucht, deren zugehöriger Objektpunkt M_5 außerhalb von π liegt. Mit der bereits bekannten Homographie kann die Abbildung des Objektpunktes M'_5 in die Bildebene von Ansicht 2 transformiert werden. Aufgrund der Tatsache, dass sich die Homographie auf die Ebene π bezieht, handelt es sich bei der Abbildung m'_{25} um die Projektion des Punktes M'_5 . M'_5 liegt in π auf dem optischen Strahl, der durch M_5 und C_1 verläuft. Aus diesem Grund wird das Verfahren als virtuelle Parallaxe bezeichnet²²⁵. Da der optische Strahl einer Epipolarlinie in Ansicht 2 entspricht, kann diese aus dem Vektorprodukt der zwei Abbildungen \tilde{m}_{25} und \tilde{m}'_{25} berechnet werden:

$$l_{25} = \tilde{m}_{25} \times \tilde{m}'_{25}$$

Formel 84

Der Prozess wird für einen weiteren, außerhalb von π liegenden Objektpunkt M_6 wiederholt. Dies liefert eine zweite Epipolarlinie, mit der der Epipol in Ansicht 2 berechnet werden kann²²⁶:

$$e_2 = l_{25} \times l_{26}$$

Formel 85

Somit stehen alle notwendigen Komponenten zur Berechnung der Fundamental-Matrix zur Verfügung. Die Homographie kann dabei helfen, die Tiefenstruktur einer Szene zu analysieren. Kann eine Szene in unterschiedliche Ebenen zerlegt werden, kann auch eine Homographie berechnet werden. Werden mehr als vier Punktkorrespondenzen gesucht, so sind diese direkt aus der Homographie ableitbar. Dies ist für Bildinhalte vorteilhaft, in denen die klassische Korrespondenzanalyse z.B. aufgrund schwach texturierter Bereiche nur unbefriedigende Punktkorrespondenzen liefert. Weiterhin spielt die Homographie eine Rolle bei der Generierung von virtuellen achsparallelen Stereoansichten, der sog. Rektifikation.

²²⁴ Vgl. Schreer 2005, S. 101.

²²⁵ Vgl. Hartley; Zisserman 2003, S. 335

²²⁶ Vgl. Schreer 2005, S. 102.

3.3.6 Rektifikation

Wie bereits geschildert, wird für Bewegtbild-Aufnahmen z.B. in virtual sets durchaus mit einem konvergenten Stereokamerasystem gearbeitet. Es wurde ebenfalls beschrieben, dass die mathematischen Beziehungen zwischen den beiden Ansichten in einem achsparallelen System weniger kompliziert sind als in einem konvergenten. Die Rektifikation dient dazu, über eine lineare Transformation die Ansichten eines beliebigen Stereokamerasystems in die Ansichten eines virtuellen achsparallelen Systems zu überführen. Dies führt dazu, dass für das sichtbare Stereobildpaar die Vorteile des konvergierenden Systems und für die mathematischen Berechnungen im Hintergrund (z.B. zum Erstellen einer Tiefenkarte) die Geometrie des achsparallelen Systems genutzt werden kann ²²⁷. Im Hinblick auf die Korrespondenzbeziehungen von Epipolarlinien und Bildpunkten kann nach einer erfolgreichen Rektifikation der Suchbereich auf die horizontale Richtung, d.h. die Bildzeilen eingeschränkt werden. Darüber hinaus können beispielsweise unerwünschte vertikale Disparitäten, Rotationsunterschiede und andere perspektivische Verzerrungen eliminiert werden. Eine zentrale Bedingung dabei ist, dass die optischen Zentren der beiden Kameras vor und nach der Rektifikation identisch sind. Aus diesem Grund handelt es sich bei der Transformationsmatrix der Rektifikation um eine Homographie. Eine weitere Forderung ist, dass beide rektifizierten Bildebenen in einer gemeinsamen fokalen Ebene liegen. Nach Schreer gliedert sich die Rektifikation in zwei Verarbeitungsschritte ²²⁸. Zuerst muss die Transformationsvorschrift bestimmt werden, nach der die Bildpunkte der Originalansicht in die neue Ansicht überführt werden, so dass korrespondierende Bildpunkte auf einer Zeile liegen und die Epipolarlinien horizontal verlaufen. Die Verfahren zur Bestimmung der Transformationsvorschrift unterscheiden sich darin, dass sie abhängig von den bekannten Parametern der Stereogeometrie eingesetzt werden, d.h. ob es sich um kalibrierte oder unkalibrierte Kameras handelt. In diesem Fall wird von unkalibrierten Kameras ausgegangen, wobei lediglich eine bestimmte Anzahl bekannter Punktkorrespondenzen in den beiden Ansichten vorliegt.

3.3.6.1 Mathematische Beziehungen

Ein Bildpunkt \mathbf{m}_{io} wird durch die Rektifikation überführt in den Bildpunkt \mathbf{m}_{ir} . Zunächst wird der Zusammenhang hergestellt zwischen der Projektion eines Objektpunkts M_w auf die originale und die rektifizierte Bildebene einer Kamera. Die allgemeine Projektionsgleichung in homogenen Koordinaten lautet ²²⁹:

$$\lambda \tilde{\mathbf{m}} = \mathbf{P} \tilde{\mathbf{M}}_w, \tilde{\mathbf{m}} = \begin{pmatrix} U \\ V \\ S \end{pmatrix}, \mathbf{m} = \begin{pmatrix} U/S \\ V/S \end{pmatrix} \text{ mit } S \neq 0$$

Formel 86

²²⁷ Vgl. Wöhler 2009, S. 60.

²²⁸ Vgl. Schreer 2005, S. 107.

²²⁹ Vgl. Schreer 2005, S. 107

Die Projektionsmatrix kann umgeformt werden, indem die ersten drei Spalten durch die Zeilenvektoren q_i^T dargestellt werden ²³⁰:

$$P = \begin{pmatrix} q_1^T & q_{14} \\ q_2^T & q_{24} \\ q_3^T & q_{34} \end{pmatrix} = (Q | \bar{q})$$

Formel 87

Für das optische Zentrum C gilt dann:

$$P (C \quad 1)^T = \mathbf{0} \text{ und } C = -Q^{-1}\bar{q}$$

Formel 88

Über die Matrix P_r wird die betreffende Ansicht rektifiziert. Bezüglich eines dreidimensionalen Objektpunkts gelten die folgenden Projektionsgleichungen für die ursprüngliche und die rektifizierte Ansicht ²³¹:

$$\lambda_o \tilde{m}_o = P_o \tilde{M}_w$$

Formel 90

$$\tilde{m}_r = H \tilde{m}_o \quad \text{mit } H = Q_r Q_o^{-1}$$

Formel 91

Die Gleichung für den optischen Strahl, welcher durch m_o verläuft, lautet ²³²:

$$M_w = C_o + \lambda_o Q_o^{-1} \tilde{m}_o$$

Formel 91

Nach Schreer ergibt sich daraus durch Umformen für die Beziehung zwischen einem Bildpunkt in der Originalansicht und seiner rektifizierten Projektion ²³³:

$$\tilde{m}_r = Q_r Q_o^{-1} \tilde{m}_o$$

Formel 92

²³⁰ Vgl. Fusiello; Trucco; Verri 2000, S. 17.

²³¹ Vgl. Schreer 2005, S. 108.

²³² Vgl. Fusiello; Trucco; Verri 2000, S. 17.

²³³ Vgl. Schreer 2005, S. 108.

Für die Homographie-Matrix zwischen der Originalansicht rektifizierter Ansicht ergibt sich daraus ²³⁴:

$$\tilde{\mathbf{m}}_r = \mathbf{H} \tilde{\mathbf{m}}_o \quad \text{mit } \mathbf{H} = \mathbf{Q}_r \mathbf{Q}_o^{-1}$$

Formel 93

3.3.6.2 Bestimmung der Transformationsvorschrift

Bei unkalibrierten Kameras sind die intrinsischen und extrinsischen Parameter unbekannt. Es wird von dem Fall ausgegangen, dass hinreichend viele und entsprechend genaue Punktkorrespondenzen vorliegen. Es gilt ²³⁵:

$$\tilde{\mathbf{m}}_{1i} \rightarrow \tilde{\mathbf{m}}_{2i}$$

Formel 94

Es können dann zwei Fälle unterschieden werden. Im ersten Fall ist die Fundamental-Matrix bekannt, da sie vor der Rektifikation über die in den vorherigen Abschnitten erläuterten Verfahren ermittelt wurde. Im zweiten Fall benötigt das eingesetzte Verfahren keine explizite Berechnung der Fundamental-Matrix.

- **Fall 1:** Fundamental-Matrix bekannt: Bei Kenntnis der Fundamental-Matrix können die Homographien \mathbf{H}_1 und \mathbf{H}_2 bestimmt werden, welche die beiden Originalansichten in rektifizierte Ansichten überführen. Bei einem achsparallelen System liegen die beiden Epipole im Unendlichen, die Epipolarlinien verlaufen parallel ²³⁶:

$$\tilde{\mathbf{e}}_1 = \tilde{\mathbf{e}}_2 = (1,0,0)^T$$

Formel 95

Die Homographie der zweiten Ansicht \mathbf{H}_2 kann bestimmt werden, indem für die Homographie-Transformation eine näherungsweise starre Transformation für einen ausgewählten Punkt genutzt wird:

$$\mathbf{H}_2 = \mathbf{G} \mathbf{R} \mathbf{T}$$

Formel 96

Über die Matrix \mathbf{T} wird der Punkt in den Ursprung verschoben, welcher sich in der Bildmitte befindet ²³⁷. Die Matrix \mathbf{R} dreht den als bekannt vorausgesetzten

²³⁴ Vgl. Fusiello; Trucco; Verri 2000, S. 18.

²³⁵ Vgl. Schreer 2005, S. 108.

²³⁶ Vgl. Hartley; Zisserman 2003, S. 304.

²³⁷ Vgl. Hartley; Zisserman 2003, S. 304.

Epipol auf die Position $(f, 0, l)^T$ auf der horizontalen Achse. Die Transformationsmatrix \mathbf{G} überführt den Epipol in eine Position im Unendlichen²³⁸.

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \\ f & & \end{bmatrix}$$

Formel 97

Daraus ergibt sich der transformierte Epipol $\mathbf{e}_l^{\sim} = (f, 0, 0)^T$. Jeder andere Bildpunkt (u, v, l) wird mit Hilfe der Transformation in einen Punkt $(u', v', l')^T = (u, v, l - u/f)^T$ überführt²³⁹. Aufgrund der o.g. genannten Forderung, dass beide rektifizierten Bildebenen in der gleichen Ebene liegen, müssen die korrespondierenden Epipolarlinien identisch sein. Aus der Homographie-Transformation für Linien ergibt sich demnach:

$$\mathbf{H}_1^{-T} \tilde{\mathbf{l}}_1 = \mathbf{H}_2^{-T} \tilde{\mathbf{l}}_2$$

Formel 98

Jedes Paar von Transformationen, für die die o.g. Beziehung erfüllt ist, wird als zusammengehöriges Paar von Transformationen bezeichnet²⁴⁰. Nachdem die Homographie-Matrix von Ansicht 2 \mathbf{H}_2 ermittelt wurde, folgt die Bestimmung von \mathbf{H}_1 . Um die Ähnlichkeit zwischen ursprünglicher und rektifizierter Ansicht zu gewährleisten, muss der mittlere quadratische Abstand der transformierten Bildpunktpositionen korrespondierender Bildpunkte möglichst gering sein. Hierfür kann folgendes Optimierungskriterium verwendet werden²⁴¹:

$$\sum_i d(\mathbf{H}_1 \tilde{\mathbf{m}}_{1i,r}, \mathbf{H}_2 \tilde{\mathbf{m}}_{2i,r})$$

Formel 99

Die projektive Transformation \mathbf{H}_1 ist nur dann eine zugehörige Transformation zu \mathbf{H}_2 , wenn gilt:

$$\mathbf{H}_1 = (\mathbf{I} + \mathbf{H}_2 \tilde{\mathbf{e}}_2 \mathbf{a}^T) \mathbf{H}_2 \mathbf{M} = (\mathbf{I} + \mathbf{H}_2 \tilde{\mathbf{e}}_2 \mathbf{a}^T) \mathbf{H}_0$$

Formel 100

²³⁸ Vgl. Hartley; Zisserman 2003, S. 305.

²³⁹ Vgl. Schreer 2005, S. 113.

²⁴⁰ Vgl. Hartley; Zisserman 2003, S. 305.

²⁴¹ Vgl. Schreer 2005, S. 114.

Die beliebige, nicht-singuläre 3×3 -Matrix \mathbf{M} ergibt sich aus der Faktorisierung von \mathbf{F} über $\mathbf{F} = [\tilde{\mathbf{e}}_2]_{n \times n}^* \mathbf{H}$. Aufgrund dessen, dass der Epipol $\tilde{\mathbf{e}}_2$ im Unendlichen liegt, resultiert daraus die Identität²⁴²:

$$(\mathbf{I} + \mathbf{H}_2 \tilde{\mathbf{e}}_2 \mathbf{a}^T) = \mathbf{I} + (1, 0, 0)^T \mathbf{a}^T$$

Formel 101

Diese muss folgende Form besitzen:

$$\mathbf{H}_A = \begin{pmatrix} a & b & c \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Formel 102

\mathbf{H}_A stellt eine affine Transformation dar, d.h. sie erhält die Kollinearität. Wenn also $\mathbf{H}_1 = \mathbf{H}_A \mathbf{H}_0$ und $\mathbf{H}_0 = \mathbf{H}_2 \mathbf{M}$ gültig ist, bilden \mathbf{H}_1 und \mathbf{H}_2 ein zusammengehöriges Transformationspaar. Für die korrespondierenden Bildpunkte in den rektifizierten Ansichten ergeben sich die Gleichungen²⁴³:

$$\tilde{\mathbf{m}}_{1i,r} = \mathbf{H}_0 \tilde{\mathbf{m}}_{1i} \text{ und } \tilde{\mathbf{m}}_{2i,r} = \mathbf{H}_2 \tilde{\mathbf{m}}_{2i}$$

Formel 103

Aufgrund der Bedingung für $\tilde{\mathbf{e}}_2$ kann das Optimierungskriterium aus Formel 99 auf die Matrix \mathbf{H}_A übertragen werden.

$$\sum_i d(\mathbf{H}_A \tilde{\mathbf{m}}_{1i,r}, \mathbf{H}_2 \tilde{\mathbf{m}}_{2i,r})$$

Formel 104

Durch Umformen ergibt sich der folgende Ausdruck²⁴⁴:

$$\sum_i d(a u_{1i,r} + b v_{1i,r} + c u_{2i})^2$$

Formel 105

Wenn \mathbf{M} aus der Faktorisierung der Fundamental-Matrix vorliegt und die

²⁴² Vgl. Hartley; Zisserman 2003, S. 306.

²⁴³ Vgl. Schreer 2005, S. 115.

²⁴⁴ Vgl. Schreer 2005, S. 115.

Homographie-Transformation H_2 bestimmt wurde, ergibt sich aus dem Optimierungskriterium ein lineares Minimierungsproblem nach dem kleinsten quadratischen Fehler. Zur Lösung können lineare Verfahren aus der mathematischen Numerik wie z.B. die Singularwertzerlegung verwendet werden ²⁴⁵. Als Ergebnis ergeben sich die Parameter a, b und c der Matrix H_A , so dass die benötigte Transformation H_1 berechnet werden kann ²⁴⁶. Dadurch sind beide Homographie-Matrizen bekannt, mit denen die Originalansichten in die rektifizierten Ansichten transformiert werden können.

• **Fall 2:** Fundamental-Matrix unbekannt: Bei diesem Fall wird die Gegebenheit genutzt, dass für ein korrespondierendes Paar von Bildpunkten die Epipolarbedingung nach der Rektifikation weiterhin gelten muss. Dabei gilt für das rektifizierte Bildpunktpaar die bekannte Fundamental-Matrix ²⁴⁷:

$$F_r = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$$

Formel 106

$$(H_2 \tilde{m}_2)^T F_r (H_1 \tilde{m}_1) = 0$$

Formel 107

Die Matrix H_2 wird auch hier nach Formel 96 über Faktorisierung ermittelt. Die Schätzung der Matrix H_1 erfolgt unter Verwendung der Epipolargleichung für die rektifizierten Ansichten (Formel 107). Für N vorliegende Punktkorrespondenzen wird folgendes Optimierungskriterium beschrieben ²⁴⁸:

$$F(H_1, H_2) = \sum_{i=1}^N ((H_2 \tilde{m}_{2i})^T F_r (H_1 \tilde{m}_{1i}))^2$$

Formel 108

Die erste Zeile der rektifizierenden Fundamental-Matrix enthält den Nullvektor, das Optimierungskriterium berücksichtigt daher lediglich die zweite und dritte Zeile der Homographie-Matrizen. Um eine Matrix H_1 mit vollem Rang zu ermitteln, können jetzt drei beliebige reelle Zahlen bestimmt werden. Um

²⁴⁵ Vgl. Schreer 2005, S. 252.

²⁴⁶ Vgl. Hartley; Zisserman 2003, S. 307.

²⁴⁷ Vgl. Faugeras; Loung 2001, S. 372.

²⁴⁸ Vgl. Schreer 2005, S. 116.

allerdings ein optimiertes Ergebnis für die Stereoanalyse zu erhalten, wird für die erste Zeile von \mathbf{H}_l ein weiteres Optimierungskriterium definiert ²⁴⁹:

$$\sum_{i=1}^N ((\mathbf{H}_1 \tilde{\mathbf{m}}_{1i})_u - (\mathbf{H}_2 \tilde{\mathbf{m}}_{2i})_u)^2$$

Formel 109

Damit wird gewährleistet, dass die horizontalen Komponenten korrespondierender Bildpunkte in den rektifizierten Ansichten möglichst gleich sind. Auf diese Weise werden Verzerrungen und zu starke Unterschiede zwischen den rektifizierten Ansichten verringert. Die fehlende Homographie-Matrix \mathbf{H}_l kann wiederum mit Hilfe linearer Verfahren nach dem kleinsten Fehlerquadrat ermittelt werden. Sind \mathbf{H}_l und \mathbf{H}_2 bekannt, können die Epipole der Originalansichten berechnet werden:

$$\mathbf{e}_i = \mathbf{H}_i^{-1}(1,0,0)^T$$

Formel 110

Die Fundamental-Matrix der Originalansicht lässt sich dann aus der Fundamental-Matrix der rektifizierten Ansicht berechnen ²⁵⁰:

$$\mathbf{F} = \mathbf{H}_1^T \mathbf{F}_r \mathbf{H}_2$$

Formel 111

3.3.6.3 Durchführung der Transformation

Mit den vorliegenden Homographie-Matrizen müssen die Pixelpositionen transformiert und für diese Positionen die zugehörigen Grau- bzw. Farbwerte ermittelt werden. Die Gleichung für die Transformation von den Koordinaten der Originalansicht in die Koordinaten der rektifizierten Ansicht lautet nach Schreer ²⁵¹:

$$\tilde{\mathbf{m}}_{ri} = \mathbf{Q}_{ri} \mathbf{Q}_{oi}^{-1} \tilde{\mathbf{m}}_{oi} = \mathbf{H}_i \tilde{\mathbf{m}}_{oi} \text{ mit } \mathbf{H}_i = \mathbf{Q}_{ri} \mathbf{Q}_{oi}^{-1} \text{ und } i = 1, 2$$

Formel 112

Da die Grauwerte für das Raster der rektifizierten Ansicht gesucht werden, wer-

²⁴⁹ Vgl. Schreer 2005, S. 116.

²⁵⁰ Vgl. Schreer 2005, S. 117.

²⁵¹ Vgl. Schreer 2005, S. 117.

den die inverse Transformation $H_1^{(-l)}$ und $H_2^{(-l)}$ diskreter Pixelkoordinaten in der rektifizierten Ansicht auf Positionen im Originalbild genutzt (Rückwärts-Transformation) ²⁵²:

$$\tilde{m}_{oi} = H_i^{-1} \tilde{m}_{ri}$$

Formel 113

Durch die Rektifikation ist es möglich, dass sich der aktive Bildbereich zwischen Originalansicht und rektifizierter Ansicht unterscheidet. Aus diesem Grund müssen die rektifizierenden Transformationen bezüglich des aktiven Bildbereichs des rektifizierten Bildes überprüft und ggf. korrigiert werden ²⁵³. Abhängig von Position und Orientierung der Bildebenen der Originalansicht kann z.B. der aktive Bildbereich der rektifizierten Kamera gegenüber der realen Kamera eine unerwünschte Verschiebung oder Rotation aufweisen, welche entsprechend ausgeglichen werden müssen.

Translation: Um die Verschiebung des aktiven Bildbereichs auszugleichen, werden die Eckkoordinaten der Originalbilder verschoben. Für eine horizontale Verschiebung gilt, dass diese in einer der beiden Ansichten erfolgen kann und unabhängig von der Translation in der zweiten Ansicht ist. Eine vertikale Verschiebung jedoch muss immer für beide Ansichten vorgenommen werden, um die identische Höhe korrespondierender Epipolarlinien zu gewährleisten. Die Verschiebungsmatrizen für die entsprechenden Eckkoordinaten lauten ²⁵⁴:

$$T_1 = \begin{bmatrix} 1 & 0 & -u_{min,1} \\ 0 & 1 & -\min(v_{min,1}, v_{min,2}) \\ 0 & 0 & 1 \end{bmatrix}$$

Formel 114

$$T_2 = \begin{bmatrix} 1 & 0 & -u_{min,2} \\ 0 & 1 & -\min(v_{min,1}, v_{min,2}) \\ 0 & 0 & 1 \end{bmatrix}$$

Formel 115

Skalierung: Beinhaltet die Rektifikation eine virtuelle Rotation, so besteht die Gefahr, dass beide Ansichten in Bereiche außerhalb des aktiven Bildbereiches gedreht werden. Hier schafft eine Skalierung Abhilfe, um die beiden

²⁵² Vgl. Fusiello; Trucco; Verri 2000, S. 18.

²⁵³ Vgl. Schreer 2005, S. 118.

²⁵⁴ Vgl. Schreer 2005, S. 118.

rektifizierten Ansichten an den aktiven Bildbereich anzupassen. Die Skalierungsmatrix \mathbf{S} ergibt sich über die Bestimmung der maximalen horizontalen und vertikalen Auflösungen bezogen auf die verfügbare Bildauflösungen in den jeweiligen Richtungen N_u und N_v ²⁵⁵:

$$s_u = \frac{\max((v_{\max,1} - u_{\min,1}), (u_{\max,2} - u_{\min,2}))}{N_u}$$

Formel 116

$$s_v = \frac{\max((v_{\max,1} - v_{\min,1}), (v_{\max,2} - v_{\min,2}))}{N_v}$$

Formel 117

$$\mathbf{S} = \begin{bmatrix} s_u & 0 & 0 \\ 0 & s_v & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Formel 118

Aus der Notwendigkeit der Translation und Skalierung ergeben sich neue Gleichungen für die Homographie-Matrizen:

$$\mathbf{H}_{i,neu} = \mathbf{S} \mathbf{T}_i \mathbf{H}_i$$

Formel 119

• Rückwärts-Transformation (backward-mapping): Wird eine Rücktransformation durchgeführt, so liegt die Pixelposition im Originalbild i.d.R. nicht auf dem Pixelraster, sondern auf einer Zwischenposition. Die Grau- bzw. Farbwerte der neuen Position können mit Hilfe einer sog. bilinearen Interpolation aus den Werten der vier umliegenden Pixel errechnet werden. Dabei ist der gesuchte Wert abhängig von der jeweiligen Entfernung der vier umliegenden Bildpunkte²⁵⁶. Die Transformation kann dargestellt werden über die Inverse $\mathbf{T}^{(-)}$ der Translationsmatrix.

²⁵⁵ Vgl. Schreer 2005, S. 120.

²⁵⁶ Vgl. Waldowski (Hg) 2010, S. 7f.

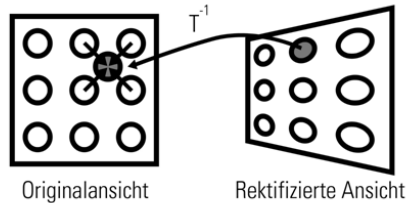


Abbildung 17: Rückwärts-Transformation

•Vorwärts-Transformation (forward-mapping): Die hier verwendete Interpolationsvorschrift wird als Nächster-Nachbar-Interpolation (nearest neighbour) bezeichnet. Sie bietet sich jedoch eher an für Anwendungen, bei denen lediglich bestimmte Bildbereiche verarbeitet werden. Dabei wird zunächst durch eine Rückwärts-Transformation dem gesuchten Pixel im rektifizierten Bild der Wert des am nächsten liegenden Pixels im Originalbild zugewiesen ²⁵⁷. Indem die Transformationsvorschrift einmalig berechnet wird, ist der Zusammenhang zwischen den Pixelpositionen in der rektifizierten Ansicht und den nächsten Bildpunkten im Originalbild festgelegt ²⁵⁸. Die Grau- bzw. Farbwerte können dann direkt aus dem Originalbild in die rektifizierte Position kopiert werden. Die jeweiligen Zuordnungen werden in einer Tabelle gespeichert, wodurch die Verarbeitungsgeschwindigkeit zur Durchführung der Rektifikation steigern lässt.

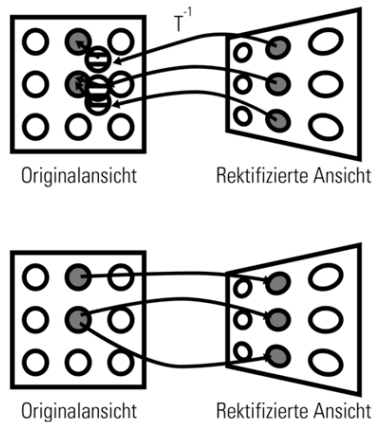


Abbildung 18: Vorwärtstransformation

²⁵⁷ Vgl. Waldowski (Hg) 2010, S. 7f.

²⁵⁸ Vgl. Schreer 2005, S. 122.

4 Zusammenfassung

Bei jeder Kameraaufnahme wird eine dreidimensionale Szene über eine Projektion in ein zweidimensionales Bild überführt. Die Projektionsart der Zentralprojektion berücksichtigt den Faktor der perspektivischen Verzerrung entsprechend der menschlichen Wahrnehmung. Mathematisch betrachtet handelt es sich bei der Projektionsbeziehung zwischen Szenenraum und Bildraum um eine räumliche Ähnlichkeitstransformation. Diese Transformation kann mit Hilfe der homogenen Koordinaten als Matrix formuliert werden. So wird der Zusammenhang hergestellt zwischen den Koordinaten des Szenenpunktes und den Koordinaten des Bildpunktes. Bei einem Stereoaufbau wird ein Szenenpunkt in zwei Bildpunkten abgebildet. Stereosysteme beinhalten zwei Kameras und werden entsprechend deren Ausrichtung als konvergent oder achsparallel bezeichnet. Die mathematischen Beziehungen zwischen den Kameras werden in der Epipolargeometrie zusammengefasst. Die Fundamental-Matrix eines Stereosystems enthält die gesamten Informationen der Epipolargeometrie und kann mit Hilfe von korrespondierenden Bildpunkten ermittelt werden. Über die Rektifikation lässt sich eine beliebige Stereoanordnung algorithmisch in ein achsparalleles System überführen.

Literaturverzeichnis

Monografien

- Ascher, Steven; Pincus, Edward (2007): The filmmaker's handbook. A comprehensive guide for the digital age. 3. ed. /// completely rev. and updated by Steven Ascher with contributions by David Leitner. New York, London /// New York: Plume; Penguin.
- Bender, Michael; Brill, Manfred (2006): Computergrafik. Ein anwendungsorientiertes Lehrbuch. 2., überarb. Aufl. München: Hanser.
- Birn, Jeremy (2007): Lighting & Rendering. [3D-Grafiken meisterhaft beleuchten; realistische Texturen entwickeln]. 2. komplett überarb. und erw. Ausg. München: Addison-Wesley.
- Böhringer, Joachim; Bühler, Peter; Schlaich, Patrick (2008): Kompendium der Mediengestaltung. [Produktion und Technik] für Digital- und Printmedien. 4., vollst. überarb. u. erw. Aufl. Berlin: Springer (Kompendium der Mediengestaltung Digital und Print).
- Burg, Klemens; Haf, Herbert; Wille, Friedrich (2008): Höhere Mathematik für Ingenieure. Band II: Lineare Algebra. 6., überarbeitete Auflage. Wiesbaden: B.G. Teubner Verlag / GWV Fachverlage GmbH Wiesbaden.
- Cheshire, David; Hermstein, Rudolf (1981): Filmen. Bern [u.a.] /// Bern: Hallwag.
- Dobbert, Tim (2005): Matchmoving. The invisible art of camera tracking. San Francisco, Calif., London: SYBEX.
- Doeffinger, Derek (2009): Creative Shutter Speed: Master the Art of Motion Capture: John Wiley and Sons Ltd.
- Faugeras, Olivier; Loung, Quang-Tuan (2001): The geometry of multiple images. The laws that govern the formation of multiple images of a scene and some of their applications. Cambridge, Mass.: MIT Press.
- Hartley, Richard; Zisserman, Andrew (2003): Multiple view geometry in computer vision. 2nd ed. Cambridge UK, New York: Cambridge University Press.
- Jähne, Bernd (2005): Digitale Bildverarbeitung. 6., überarbeitete und erweiterte Auflage. Berlin, Heidelberg: Springer.
- Lipton, Lenny (1982): Foundations of the stereoscopic cinema. A study in depth. New York: Van Nostrand Reinhold.
- Mahler, Gerhard (2005): Die Grundlagen der Fernsehtechnik. Systemtheorie und Technik der Bildübertragung. Berlin, Heidelberg: Springer.
- Meschede, Dieter (2008): Optik, Licht und Laser. 3., durchges. Aufl. Wiesbaden: Vieweg + Teubner.
- Neumann, Burkhard (2005): Bildverarbeitung für Einsteiger. Programmbeispiele mit Mathcad. Berlin Heidelberg: Springer
- Poynton, Charles A. (2003): Digital video and HDTV. Algorithms and interfaces. Amsterdam, Boston: Morgan Kaufmann Publishers.
- Schmidt, Ulrich (2009): Professionelle Videotechnik. Grundlagen, Filmtechnik,

Fernsehtechnik, Geräte- und Studioteknik in SD, HD, DI, 3D. 5., aktualisierte und erweiterte Auflage. Berlin, Heidelberg: Springer

- Schreer, Oliver (2005): Stereoanalyse und Bildsynthese. Berlin, Heidelberg: Springer.
- Schröder, Gottfried; Treiber, Hanskarl (2007): Technische Optik. Grundlagen und Anwendungen. 10., erw. Aufl. Würzburg: Vogel (Vogel-Fachbuch).
- Steinmüller, Johannes (2008): Bildanalyse. Von der Bildverarbeitung zur räumlichen Interpretation von Bildern. Berlin /// Berlin, Heidelberg: Springer (eXamen).
- Wöhler, Christian (2009): 3D Computer Vision. Efficient Methods and Applications. Berlin: Springer (X).

Fachartikel

- Faugeras, Olivier et al. (1993): Real time correlation-based stereo: algorithm, implementations and applications. Unter Mitarbeit von Bernard Hotz, Herve Mathieu, Thierry Vieville et al. Institut National de Recherche en Informatique et en Automatique (INRIA).
- Fischler, Martin A.; Bolles, Robert C. (1981): Random Sample Consensus. A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. In: Communications of the ACM, Jg. 24, H. 6, S. 381–395.
- Fusiello, Andrea; Trucco, Emanuele; Verri, Alessandro (2000): A compact algorithm for rectification of stereo pairs. In: Machine vision and applications, Vol. 12, No. 1 (2000), S. 16–22.
- Fusiello, A. Irsara, L. (2008): Quasi-Euclidean Uncalibrated Epipolar Rectification. Unter Mitarbeit von L. Irsara. University of Verona (IT), Dipartimento di Informatica.
- Litwiller, D. (2001): CCD vs. CMOS Facts and Fiction. In: Photonics spectra, Vol. 35, No. 1 (2001), S. 154–158.
- Litwiller, D. (2005): CMOS vs. CCD Maturing Technologies, Maturing Markets. In: Photonics spectra, Vol. 39, No. 8 (2005), 54–61.
- Sur, Frédéric; Noury, Nicolas; Berger, Marie-Odile (2008): Computing the Uncertainty of the 8 Point Algorithm for Fundamental Matrix Estimation.
- Zhang, Z. (1998): Determining the Epipolar Geometry and its Uncertainty: A Review. In: International journal of computer vision, Vol. 27, No. 2 (1998), S. 161–195.

Hochschulschriften

- Eschenburg, Jonas (2006): Optisches Kameratracking anhand natürlicher Merkmale. Diplomarbeit. Augsburg. Universität Augsburg, Institut für Informatik.
- Hofmeyer, Frank (2008): Analyse und prototypische Umsetzung eines Prozesses zur stereoskopischen Fusion von CG-Environments und Real-Footage-Elementen. Masterthesis. Betreut von Nikolaus Hottong und Michael Waldowski. Furtwangen. Hochschule Furtwangen University, Fakultät Digitale Medien.
- Hottong, Nikolaus (Hg.) (2009): Stereoskope HD-Produktion. Grundlagen Kon-

zepte Testergebnisse. Furtwangen: Fakultät Digitale Medien Hochschule Furtwangen (Schriftenreihe der Fakultät Digitale Medien an der Hochschule Furtwangen).

Internetquellen

- Cohen, David S. (2008): James Cameron supercharges 3-D. 'Avatar' helmer reveals the art & science of stereo. Herausgegeben von Variety. Online verfügbar unter <http://www.variety.com/article/VR1117983864.html>, zuletzt geprüft am 25.05.2010.
- RED Digital Cinema Camera Company (Hg.): RED ONE Tech Specs. Online verfügbar unter http://www.red.com/cameras/tech_specs/, zuletzt geprüft am 25.05.2010.
- Seymour, Mike (2008): Art of Digital 3D Stereoscopic Film. Online verfügbar unter <http://www.fxguide.com/article471.html>.

Sonstige Veröffentlichungen

- Puts, Erwin (2002): Leica M Objektive. Ihre Seele und ihre Geheimnisse. Herausgegeben von Leica Camera AG.
- Waldowski, Michael (Hg.) (2010): Skript Grafische Datenverarbeitung. Unter Mitarbeit von André Bohnet-Armstrong, Franziska Barth und Kathrin Kühnle et al. Hochschule Furtwangen University.
- Stereoscopic Technologies GmbH (Hg.) (2010): Product Catalog 2010.

Teil 4

Stereoskopische Fusion von CG-Environments und Real-Footage Elementen

Frank Hofmeyer

Abbildungsverzeichnis

Abbildung. 1.3: Darstellung der Komplementärfarben Blau und Orange	205
Abbildung. 1.4: Original U.S. Patent "Method of Producing Moving Picture Cartoons"	207
Abbildung. 2.1: Erster Storyboardauszug mit integriertem Scheinfenster	211
Abbildung. 2.2: Zweiter Storyboardauszug mit integriertem Scheinfenster	211
Abbildung. 2.3: Gesamtansicht der originalen Aufnahme	215
Abbildung. 2.4: Erstellung einer Garbage Matte	216
Abbildung. 2.5: Die noch unbearbeitete Status-Ansicht	218
Abbildung. 2.6: Das finale Chroma-Keying mit Keylight	218
Abbildung. 2.7: Detailansicht des Chroma-Keyings	219
Abbildung. 2.8: Fehlerhafte Maske nach dem Chroma-Keying	220
Abbildung. 2.9: Korrigierte Maske nach dem Rotoscoping	220
Abbildung. 2.10: Die Zentralprojektion	222
Abbildung. 2.11: Das Sichtvolumen bei der Zentralprojektion	223
Abbildung. 2.12: Der Öffnungswinkel bei der Zentralprojektion	224
Abbildung. 2.13: Die 3D-Szene ohne Texturen	226
Abbildung. 2.14: Die beleuchtete 3D-Szene mit Texturen	227
Abbildung. 2.15: Übersicht über die gesamte Szene im Maya	228
Abbildung. 2.17: Ein Einzelbild der fertigen Animation aus der linken Ansicht	231
Abbildung. 2.18: Der 3D-Brillen-Effekt in After Effects	232
Abbildung. 2.19: Qualitätsverlust durch das Zeilensprungverfahren	236
Abbildung. 2.20: Das Stereo-3D-PlanungsTool	241
Abbildung. 3.1: Polygonplane als Träger der Darstellertextur	243
Abbildung. 3.2: Darstellertextur mit Alphakanal auf der Plane	244
Abbildung. 3.3: Angeschnittener Darsteller während der Animation	245
Abbildung. 3.4: Ausrichten der Beine des Darstellers	245
Abbildung. 3.5: Interaktion mit Beispielerendering	246
Abbildung. 3.6: Ein Fuss durchdringt die Bodengeometrie	247
Abbildung. 3.7: Steinsplitter durchdringen die Bodengeometrie	248
Abbildung. 3.8: Billboarddarstellung von Pflanzen	250
Abbildung. 3.9: Billboarddarstellung der Pflanzen von oben	251
Abbildung. 3.10: Erweitertes Billboard einer Pflanze	252
Abbildung. 3.11: Aufbau der 2.5D-Szene entlang der z-Achse	253
Abbildung. 3.12: Stereoskopisches Ergebnis des 2.5D-Fusionsverfahrens	254
Abbildung. 3.13: Verdeckung als positiver Effekt des 2.5D-Fusionsverfahrens	255
Abbildung. 3.14: Cardboard-look des 2.5D-Fusionsverfahrens	256
Abbildung. 3.15: Sichtbare Ebenengrenzung im 2.5D-Raum	256

1 Einleitung

In diesem Artikel soll zunächst die Frage beantwortet werden, welche konventionellen Postproduktionstechniken zur Fusion von real gedrehten Schauspieler:innen mit computer-generierten Filmsets erforderlich sind. Im Vorfeld werden zunächst die verschiedenen Bezeichnungen geordnet, die international im Bereich der Filmeffekte gebräuchlich sind. Anschließend werden einige Überlegungen zu generellen Qualitätskriterien bezüglich digitaler Filmeffekte angestellt. Die folgenden Abschnitte beinhalten schließlich kurze Darstellungen der konkret benötigten Techniken Compositing, Chroma-Keying und Rotoscoping. Danach wird eine Übertragung auf den stereoskopischen Postproduktionsprozess vorgenommen unter dem Aspekt der optimalen Verschmelzung von 3D-Realfilmmaterial mit stereoskopischen-CGI-Elementen.

Alle Ausführungen basieren auf den vorangegangenen Veröffentlichungen im Arbeitspapier Nr. 5 der Fakultät Digitale Medien.

1.1 Postproduktion und Visual Effects

Unter Visual Effects (VFX), werden in Deutschland i.d.R. digitale Effekte verstanden, während der Begriff Special Effects (SFX) meist als übergeordneter Begriff verwendet wird. In den USA gilt VFX jedoch als ein Überbegriff für alle mit den Augen wahrnehmbaren Effekte. Als SFX werden alle Effekte bezeichnet, die von der Kamera live am Set aufgezeichnet werden, wie bspw. Miniaturmodelle, Make-up, Pyrotechnik, usw. Mit dem zunehmenden Einzug der digitalen Computertechnik in die Effekterstellung, wurde auch die neue Effektkategorie Digital Effects (DFX) eingeführt¹. Im Kontext dieser Arbeit wird immer von Digital Effects gesprochen werden. Grundsätzlich betrachtet existieren zwei verschiedene Ansatzpunkte für einen gelungenen digitalen Filmeffekt, nämlich die dramaturgische Kohärenz - hier im Sinne der Nachvollziehbarkeit gemeint - und die technische Ausführung an sich. Beide müssen sowohl für sich als auch untereinander stimmig sein, damit der Effekt gut funktionieren und wirken kann. Bei den dramaturgischen Gesichtspunkten ist entscheidend, ob

1 Vgl. Bertram, Sacha 2005: VFX, S. 25-26; Knoll, John 2005: Star Wars, S.2.

der Effekt im Kontext des Films glaubhaft ist und ob er die Geschichte voran bringt, sie unterstützt oder gar verbessert - oder ob er nur um des spektakulären Effekts Willen eingesetzt wird. Bei der technischen Ausführung kommt es primär darauf an, dass der Effekt vom Zuschauer nicht wahrgenommen wird². Ein DFX sollte nur dann als solcher zu erkennen sein, wenn der Zuschauer weiß, dass es sich logischerweise um einen Effekt handeln muss. Übereinstimmend dazu die Philosophie der Firma Matte World Digital³: „The Matte World Digital philosophy is to produce seamless and realistic visual effects so the audience is not aware they are viewing an effect, but rather an integral part of the director's vision. To say you haven't seen our work is our best compliment.“⁴

1.2 Compositing

Eine Arbeit, die sich mit der Fusion von CG-Environments und Real-Footage auseinandersetzt, wird unweigerlich mit dem Compositing konfrontiert, da das finale Bild letztendlich aus computergenerierten und real gedrehten 2D-Bildelementen besteht. Ein Compositing wird bei jedem VFX-Dreh benötigt, da erst in diesem Arbeitsschritt alle separat erstellten Bildteile zusammengefügt werden.

Der Begriff Compositing kommt vom englischen Verb „to compose“, was im vorliegenden Zusammenhang zusammensetzen bedeutet und beschreibt den Vorgang der Anfertigung eines Composite. Ein Composite (dt.: Verbund, Gemisch) ist „etwas, das aus verschiedenen Teilen zu einem Ganzen zusammengefügt wird.“⁵

Das digitale Compositing umfaßt die Manipulation und das Zusammenfügen von 2D-Bildelementen zu einem stimmigen, natürlichen und homogenen Gesamtbild. Das oberste Ziel beim Compositing ist es, den Zuschauer nicht erkennen zu lassen, dass das betrachtete Composite aus mehreren Einzelementen besteht!

Das Prinzip des digitalen Compositing besteht darin, die einzelnen 2D-Bildelemente in Ebenen (engl.: layers) übereinander zu schichten und dabei bestimmte Bildteile mittels Masken (engl.: mattes) transparent zu machen, so dass Bildteile auf tiefer liegenden Ebenen sichtbar werden. In der vorliegenden Betrachtung wird auf ein typisches Anwendungsbeispiel des Compositings eingegangen, nämlich der Integration eines Schauspielers (Vordergrund) in eine separat erstellte Umgebung (Hintergrund).

2 Vgl. Giesen, Rolf 2000: Künstliche Welten S. 43; Bertram, Sacha 2005: VFX, S. 24.

3 Matte World Digital ist eine auf Digital Visual Effects spezialisierte Firma in Novato (USA). Gegründet 1989 vom Visual Effects Supervisor Craig Barron. Die Firma erlangte weltweite Anerkennung durch ihre Mitarbeit an folgenden Kinoproduktionen (Auswahl): Terminator II (1991), Independence Day (1996), Armageddon (1998), Mission: Impossible 2 (2000), Jurassic Park III (2001), The Last Samurai (2003). Kontakt: Matte World Digital, 24 Digital Drive #6, Novato, California 94949. Vgl. Matteworld 2008b.

4 Matteworld 2008a, <http://www.matteworld.com/film/index.html>.

5 Bertram, Sacha 2005: VFX, S. 71.

Das finale Composite ist letztendlich die additive Bildwirkung dieser einzelnen Bildebenen.

Um die gewünschten Masken zu generieren, gibt es prinzipiell zwei Möglichkeiten: die Maske kann entweder manuell mittels Rotoscoping angefertigt werden, oder aber von einer Software per Keying automatisch generiert werden.

Damit die dreidimensionalen Szenen einer 3D-Software in das Compositing integriert werden können, müssen zweidimensionale Einzelbilder aus der Sicht einer virtuellen Kamera gerendert⁶ werden. Beim Renderingprozess können problemlos alle benötigten Masken automatisch mitberechnet und ausgegeben werden. Dies geschieht mittels schwarzweiß Bildern, bei denen alle weißen Bildteile später transparent dargestellt werden. Darüber hinaus kann jedem Pixel mittels einer Z-Maske eine Tiefeninformation, also dessen Entfernung zur virtuellen Kamera mitgegeben werden⁷.

1.3 Chroma-Keying

Keying bezeichnet einen Vorgang, bei dem eine Maske von einer Software automatisch generiert wird.¹ Benutzt die Software zum Erzeugen der Maske einen bestimmten Farbton im Bild, spricht man von einem Chroma-Key. Chroma-Keying setzt i.d.R. voraus, dass der finale Vordergrund des Compositings vor einem Greenscreen oder Bluescreen gedreht wurde. Die Software kann dann den Grün- oder Blauton extrahieren und transparent darstellen, wodurch im Compositing anstatt der ehemals grünen bzw. blauen Bildteile die darunter liegende Ebene (der Hintergrund) zu sehen ist. Die Bezeichnung Keying leitet sich dementsprechend davon ab, „dass bei diesem Vorgang Bildteile in andere Bilder eingesetzt werden, wie ein Schlüssel (engl.: key) in ein Schloss gesteckt wird.“

1.3.1 Qualitätskriterien

Eine Erscheinung, die als wichtigstes Kriterium zur Beurteilung eines Keyings gilt, sind die so genannten Matte Lines. Diese „Maskenlinien“ entstehen durch ein mangelhaft ausgeführtes Keying und stellen sich dem Zuschauer im finalen Compositing als helle oder dunkle Linie um ein Objekt im Vordergrund - bspw. um einen Schauspieler - dar, wodurch die Integration des besagten Objekts in den Hintergrund nicht mehr gelingt.

In den Anfängen des Keyings waren Matte Lines kaum zu umgehen, da für den Ef-

6 Rendering (dt. etwa Bildsynthese) bezeichnet die Erstellung zweidimensionaler Bilder aus dreidimensionalen Szenenbeschreibungen vom Betrachtungspunkt einer im Computer simulierten Kamera aus. Eine Szene zu rendern bedeutet also, aus dem Blickwinkel der Kamera auf die erstellte Szene zu schauen, das sichtbare Blickfeld zu berechnen und das Ergebnis in einer Bilddatei abzuspeichern. Vgl. Driemeyer, Thomas & Herken, Rolf 2005: Rendering with mental ray, mental ray Handbooks Vol. 1, S. 5.

7 Vgl. Bertram, Sacha 2005: VFX, S. 78.

fekt mehrere analoge Kopier- und Filtervorgänge nötig waren⁸. Aufgrund der heutigen Möglichkeiten in der digitalen Postproduktion und den ausgesprochen präzisen Einstellmöglichkeiten der Keyingsysteme, sind Matte Lines bei Film- und TV-Produktionen inzwischen nicht mehr zu tolerieren. Dasselbe gilt auch für das sog. Color Spill⁹, eine Farbüberstrahlung eines Blue oder Green Screens auf helle Objekte, die gekeyt werden sollen.

1.3.2 Welche Keying-Farbe ist am effektivsten?

Um die Frage zu klären, welche Keying-Farbe am besten geeignet ist, müssen neben Kriterien der menschlichen Farbwahrnehmung auch (kamera-)technische Gesichtspunkte in Betracht gezogen werden.

Aufgrund der Tatsache, dass beim Dreh mit Schauspielern meistens die Haut zu sehen ist, würde sich für den Keying-Prozess prinzipiell ein Hintergrund anbieten, dessen Farbe sich nicht mit dem Hautton mischen lässt - wodurch Maskenkanten deutlich besser zu lokalisieren wären.

Bereits K. HERING¹⁰ ging 1878 bei seiner Gegenfarbtheorie, einer Theorie zur Farbwahrnehmung, davon aus, dass es keine Mischfarben gibt, die als gelbblau oder als rotgrün beschrieben werden können. Er folgerte daraus, dass sich Gelb und Blau respektive Grün und Rot bei der menschlichen Farbwahrnehmung gegenseitig ausschließen. Ein gelb-oranger Hautton hebt sich folglich von einem blauen Hintergrund optimal ab. Konform hierzu wird sowohl auf dem harmonischen Farbkreis nach J. ITTEN¹¹ als auch auf der Farbsonne von H. KÜPPERS¹² die Farbe Blau als eine Komplementärfarbe zu Orange dargestellt. Dies bedeutet, dass sich Blau im Farbkreis gegenüber von Orange befindet und demzufolge die größte mögliche Distanz und den maximalen Kontrast zu Orange hat (Vgl. Abbildung 1.3). Aus diesem Grund gibt es keine Überschneidungen mit der Keyfarbe, weshalb Blau zunächst als die naheliegende Wahl für die Hintergrundfarbe beim Keying erscheint.

8 Vgl. Mulack, Thomas & Giesen, Rolf 2002: Special visual effects, S. 34-36

9 Spill steht im Englischen für überlaufen oder überquellen.

10 Hering, Karl Ewald Konstantin, * 1834 in Altgersdorf, Deutschland; † 1918 in Leipzig, Deutschland. Physiologe und Hirnforscher. Vgl. Planck2008: Max Planck Institute for the History of Science.

11 Johannes Itten, * 1888 in Süderen-Linden/Schweiz; † 1967 in Zürich, Maler, Grafiker und Kunstpädagoge, bedeutende Figur des frühen Bauhauses. Vgl Bauhaus 2008.

12 Harald Küppers, * 1928 in Müden, Deutschland. Dipl.-Ing. der Reproduktions- und Drucktechnik. Vgl. Küppers 2001: Biographie; Küppers 2005: Farbenlehre.



Abbildung. 1.3: Darstellung der Komplementärfarben Blau und Orange

Gegen Blau spricht allerdings, dass alle Videokameras bei blauem Licht sehr rausch-anfällig sind, da die Empfindlichkeit der Chips bei blauem Licht am geringsten ist¹³. Diese Tatsache lässt sich leicht mit der Blue-only-Funktion von Videomonitoren zeigen. Diese Funktion stellt den isolierten Blaukanal des Signals als Schwarz-Weiß-Bild dar, wodurch das Rauschen und Kompressions-Artefakte deutlich zu Tage treten¹⁴. Das Rauschen wird vor allem bei der tapebasierten Aufzeichnung zu einem Problem, da die Camcorder in diesem Fall anstelle der höherwertigen RGB-Signale nur YUV-Signale aufzeichnen, bei denen der Blaukanal lediglich 50% Auflösung hat. Wird außerdem noch mit einem DV-Codec aufgezeichnet, wird die Blauauflösung noch weiter reduziert und zudem noch komprimiert, was die Rauschproblematik noch verstärkt. „Etliche Postproduktionshäuser bestehen deshalb auf „Discrecording“¹⁵, da hierbei die Signale direkt vom Kamerakopf in 10 Bit YUV oder sogar in 10 Bit RGB 4:4:4 unkomprimiert auf Festplatten aufgezeichnet werden können.

Demgegenüber bietet die Farbe Grün den entscheidenden Vorteil, dass Videokameras i.d.R. eine eindeutig höhere Empfindlichkeit für grünes Licht haben und der Grünkanal immer mit 100% codiert wird. Ein weiterer Grund für die Verwendung von Grün als Keying-Farbe ist die Tatsache, dass grüne Wandfarben das Licht oft besser reflektieren als blaue, was die Ausleuchtung des Hintergrunds erleichtert.

Letztendlich sollte bei der Entscheidung für eine Screenfarbe aber auch der Farbton

¹³ Diese Aussage wurde mir von Michael Radeck per E-mail bestätigt. M. Radeck ist seit 1993 Cutter BFS (Bundesverband Filmschnitt Editor e.V., <http://www.bfs-cutter.de/>), Visual-Effect-Artist und seit Juni 2001 verantwortlicher Avid Certified Support Representative für AVID Postproduktionssysteme im Raum München/Bayern. Vgl. Radeck, Michael 2003: Biografie

¹⁴ Vgl. Radeck, Michael 2004: Artikel Keying am AVID DS Nitris.

¹⁵ Radeck, Michael 2004: Artikel Keying am AVID DS Nitris, S.157.

des später einzukopierenden Hintergrundes berücksichtigt werden. So bietet sich für eine typische Raumschiffschlacht im All der Dreh der Raumschiffe vor einem Bluescreen an, da der bläuliche Color Spill ideal mit dem späteren Hintergrund korrespondiert. Dasselbe gilt für den Fall, dass Schauspieler in einen Hintergrund aus Eis und Schnee integriert werden müssen. Soll die fertige Szene aber beispielsweise in einer Waldkulisse spielen, bietet sich die Verwendung eines Greenscreens an.

1.3.3 Die Reinheit der Keying-Farbe

Nachdem man sich grundsätzlich für die Keying-Farbe Blau oder Grün entschieden hat, muss man sich noch über die Reinheit der Farbe Gedanken machen. Prinzipiell gilt: je weniger Farbtönen sich auf dem Screen ergeben, desto besser wird das Keying-Ergebnis. Farbtönen ergeben sich z.B. durch eine unregelmäßige Ausleuchtung oder durch Schattenwurf aus dem Vordergrund.

Als Hintergrundfarben bieten sich die CIE Primaries an, die bereits 1931 mittels der CIE-Normfarbtafel¹⁶, einem wahrnehmungsbezogenen Farbbeschreibungssystem festgelegt wurden. Diese drei Norm-Spektralfarben, auch Primärfarben oder Primärreize genannt, sind monochromatische Strahlungen mit den Wellenlängen 700,0 nm für Rot, 546,1 nm für Grün und 435,8 nm für Blau. Keine dieser drei Primärfarben ist aus den beiden anderen ermisierbar¹⁷.

1.4 Rotoscoping

Unter Rotoscoping versteht man im DFX-Bereich prinzipiell das manuelle Anlegen einer Maske¹⁸. Das Wort Rotoscoping stammt jedoch ursprünglich aus der Trickfilm-animation und geht auf die im Jahr 1917 patentierte Erfindung des Rotoskops¹⁹ durch Max Fleischer²⁰ zurück (Vgl. Abbildung 1.4).

Das Rotoskop, erfunden 1915²¹, ermöglichte es erstmals, die Bewegungen eines Schauspielers aus einer Realfilm-Szene in eine gezeichnete Animationssequenz zu

16 CIE: Commission Internationale de l'Eclairage. Vgl. CIE 2007: Online verfügbar unter http://www.cie.co.at/index_je.html.

17 Vgl. Mäusl, Rudolf 1992: Repetitorium - Fernsehtechnik, S. 10; Poynton, Charles A. 2003: Digital Video And HDTV, S. 236.

18 Vgl. Mulack, Thomas & Giesen, Rolf 2002: Special visual effects, S. 40-43; Allen, Damian 2007: Encyclopedia of visual effects, S.25.

19 Vgl. Giesen, Rolf 2003: Lexikon des Trick- und Animationsfilms, S.168 und 361. Das Original U.S. Patent 1,242,674, "Method of Producing Moving Picture Cartoons" ist auf der beigelegten DVD enthalten.

20 Max Fleischer, eigentlich Maximilian Fleischer, * 19. Juli 1883 in Wien; † 11. September 1972 in Los Angeles. US-amerikanischer Cartoonist und Trickfilmproduzent. Fleischer beherrschte bis zu Walt Disneys Erfolgen in den 1930er Jahren den Trickfilmmarkt der USA. In seinem Studio entstanden unter anderem Betty Boop und Popeye the Sailor. Vgl. Giesen, Rolf 2003: Lexikon des Trick- und Animationsfilms; Wikipedia 2008c: Max Fleischer.

21 Vgl. Hagler, Jürgen 2007: Geschichte der Animation - Die wichtigsten Entwicklungsschritte ab 1900; Cohen 2000: Milestones Of The Animation Industry In The 20th Century.

übertragen. Dazu wurde die Realaufnahme mittels eines Umlenkspiegels Bild für Bild von hinten auf eine Mattscheibe im Zeichentisch projiziert. Die Haltungen des Schauspielers konnten auf diesem Weg bildgenau vom Animator nachgezeichnet und somit auf die zu animierende Figur übertragen werden²².

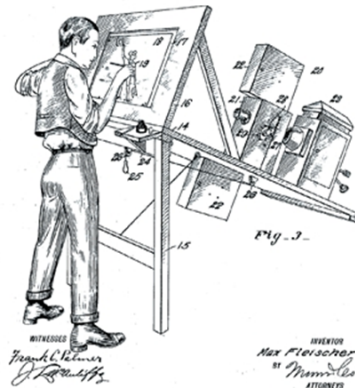


Abbildung. 1.4: Original U.S. Patent "Method of Producing Moving Picture Cartoons"

Mit diesem Verfahren, das erstmals 1914 für die Animationsserie „Out of the Inkwell“ eingesetzt wurde, war Fleischer in relativ kurzer Zeit in der Lage, seine Trickfilmfiguren mit realistischen Bewegungsabläufen zu versehen.

Obwohl die Rotoskopietechnik von einigen Puristen der Branche denunziert und abgelehnt wurde²³, bot sich das Verfahren vor allem bei komplexen Bewegungsabläufen realistisch gestalteter menschlicher Figuren an, weshalb es sich im Laufe der Zeit mehr und mehr durchsetzte. Als Paradebeispiel für den frühen Einsatz der Rotoskopie gelten die Tanzszenen in „Snow White and the Seven Dwarfs“ (dt.: "Schneewittchen und die sieben Zwerge") von 1937, dem ersten abendfüllenden Zeichentrickfilm von Walt Disney²⁴. Die aktuellste Entwicklung auf dem Gebiet der Rotoskopie stammt zweifelsfrei von Bob Sabiston, einem amerikanischen Art Director und Programmierer. Er entwickelte bereits 1997 die proprietäre Software Rotoshop, die ein interpolierendes Rotoskopieren ermöglicht. Dadurch ist es nicht mehr notwendig, jedes Einzelbild vollständig zu bearbeiten.

Statt dessen werden nur noch wichtige Schlüsselbilder von Hand mit einem Wacom Grafiktablett bearbeitet, während die dazwischenliegenden Bilder von der Software

22 Vgl. Hagler, Jürgen 2007: Geschichte der Animation - Die wichtigsten Entwicklungsschritte ab 1900; Markstein 2008b: Rotoscope.

23 Vgl. Markstein 2008: Max Fleischer Studio - Don Markstein's Toonopedia.

24 Vgl. Giesen, Rolf 2003: Lexikon des Trick- und Animationsfilms; Wikipedia 2008d: Rotoskopie.

berechnet werden. Aufgrund dieser Vorgehensweise werden wesentlich weichere Übergänge zwischen den Einzelbildern erreicht als bisher²⁵. Daraus resultiert eine deutlich lebensechtere Animation, v.a. in Bezug auf Mimik und Gestik²⁶.

Außerhalb des Vokabulars der gezeichneten Animationsfilme steht die Rotoskopietechnik, wie Eingangs erwähnt, für das Erstellen von Masken für das Compositing. Um beispielsweise einen Schauspieler aus Film A vor einem beliebigen neuen Hintergrund aus Film B zu platzieren, wurden vor dem Zeitalter der digitalen Postproduktion die Einzelbilder von Film A mit dem Rotoskop auf Papier projiziert, wodurch die Umrisse des Schauspielers exakt abgezeichnet werden konnten. Anschließend wurde der nicht benötigte Hintergrund deckend schwarz ausgemalt und das gesamte Bild abfotografiert. Auf diesem Wege erhielt man einen Maskenfilm, auf dessen Negativ der Schauspieler opak schwarz ist, während er auf dem Positivfilm vollkommen transparent abgebildet wird.

Um nun die beiden Bildteile zu kombinieren, wird zunächst nur der mit der Positivfilmmaske abgedeckte Film A auf einen neuen Film C projiziert. Folglich wird bei diesem Vorgang nur der isolierte Schauspieler übertragen. In einem zweiten Schritt wird Film C zurückgespult und der mit der Negativfilmmaske abgedeckte Film B aufkopiert, wodurch der neue Hintergrund mit einer Aussparung für den Schauspieler in das finale Bild einbelichtet wird²⁷. Bei mindestens 24 Bildern pro Sekunde ist dies ein recht aufwändiger und damit zeitraubender Prozess. Hinzu kommt noch eine gewisse Fehleranfälligkeit durch die mehrfachen Kopiervorgänge.

Der Einzug der Digitaltechnik in die Postproduktion hat das Verfahren vereinfacht. Zum einen kann der Maskenumriss nun direkt in das digitalisierte Filmmaterial eingezeichnet und beliebig oft angepasst werden und zum anderen entfallen die mehrfachen Kopiervorgänge.

Dennoch bleibt das Rotoscoping ein ausgesprochen zeit- und arbeitsintensiver Prozess. Das Haupteinsatzgebiet dieser Technik liegt heute vor allem bei der Unterstützung des Chroma-Keyings, wenn z.B. die Ausleuchtung der Greenbox zu ungleichmäßig war.

25 Vgl. Sabiston, Bob 2001: Rotoshop Animation; Flat BlackFilms 2005: Rotoshop; Wikipedia 2007: Artikel Bob Sabiston; Wikipedia 2008f: Artikel Rotoshop.

26 Das Potential dieser Technik zeigt der 2006 erschienene Kinofilm „A Scanner Darkly - Der dunkle Schirm“ von Richard Linklater. Der Film wurde u.a. mit Keanu Reeves und Robert Downey Jr. in Austin, Texas gedreht. Die Realfilmbilder wurden anschließend von Bob Sabistons Produktionsfirma Flat BlackFilms mit dem inzwischen weiterentwickelten Rotoshop bearbeitet, so dass sie letztendlich wie die Bilder eines Animationsfilms wirken. Vgl. Warner 2008: Warner Independent Pictures' A Scanner Darkly -- The Official Film site; Wikipedia 2008e: A Scanner Darkly.

27 Vgl. Bertram, Sacha 2005: VFX, S.89.

2 Stereoskopische Fusion in 2D

In diesem Kapitel wird nun erstmals die Frage nach einem möglichen Verfahren zur stereoskopischen Fusion eines realen Darstellers mit einer computergenerierten Umgebung beantwortet. Dazu wurde zunächst ein Fusionsverfahren auf Basis des konventionellen digitalen Compositings im 2D-Raum ausgewählt. Dieses erste Verfahren wurde dann mit den an der Hochschule Furtwangen zur Verfügung stehenden Mitteln im Rahmen einer Testproduktion erprobt. Die Ziele dieser Testproduktion waren:

- Überprüfung der bisherigen theoretischen Erkenntnisse auf praktische Konsistenz.
- Frühzeitige Identifikation von Problemen beim Dreh.
- Erstellung der Realaufnahmen des Darstellers als Grundlage für weitere Versuche.
- Überprüfung der prinzipiellen Machbarkeit des gewählten 2D-Fusionsverfahrens.
- Identifizierung der Vor- und Nachteile des gewählten 2D-Fusionsverfahrens.

Die praktischen Erkenntnisse aus dieser Testproduktion bildeten anschließend den Ausgangspunkt und die Vergleichsbasis für die Untersuchung weiterer alternativer stereoskopischer Fusionsverfahren (siehe Kapitel 3).

Zur Überprüfung der Durchführbarkeit der Fusionsverfahren wurde folgende Workbench verwendet: Standard PC unter Windows XP mit folgender Ausstattung:

- Adobe® After Effects® (CS3 Professional)²⁸ zur Synchronisierung der Videospuren und für das 2D-Compositing.
- Adobe® Photoshop® (CS3)²⁹ zur Erstellung des Storyboards und für das Rotoscoping.
- Ein Intuos3® Grafiktablett (A5 Wide) der Firma Wacom® Co. Ltd.³⁰ in Verbindung mit Adobe® Photoshop® zur Erstellung des Storyboards und für das Rotoscoping.

28 Adobe® After Effects® gilt als einer der Industriestandards in den Bereichen Motion Graphics und Visual Effects. Nähere Produktinformationen sind online verfügbar: <http://www.adobe.com/de/products/aftereffects/features/> Die Software wird von der Firma Adobe® entwickelt. Weitere Informationen zum Firmenprofil unter: <http://www.adobe.com/de/aboutadobe/>.

29 Adobe® Photoshop® ist der professionelle Standard zur Bildbearbeitung und -optimierung. Produktinformationen sind online verfügbar: <http://www.adobe.com/de/products/photoshop/photoshop/>.

30 Die Firma Wacom® Co. Ltd. wurde 1983 in Japan gegründet und hat heute Niederlassungen in den USA und Europa. Der Name des Unternehmens ist aus dem japanischen Wort für Harmonie wa und der Abkürzung für Computer com zusammengesetzt. Wacom entwickelte die ersten Grafiktablets mit druckempfindlichen Stift ohne Batterie und Kabel. Informationen zu den Produkten gibt es online unter: <http://www.wacom-europe.com>.

- Autodesk® Maya® (8.5)³¹ für 3D-Modellierung, Animation, Rendering und die Produktionsplanung mittels Animatics.
- Microsoft® Excel®³² zur Berechnung der verfügbaren Tiefenbereiche in der Szene.

2.1 Planung und Vorbereitung

Da es in der zur Verfügung stehenden Zeit nicht möglich war, einen stimmigen Kurzfilm zu produzieren und dies ohnehin nicht Hauptbestandteil der vorliegenden Arbeit war, fiel relativ schnell die Entscheidung, lediglich einzelne, kurze Einstellungen zu drehen. Diese waren in gleichem Maße geeignet, die geplanten Tests durchzuführen - ohne jedoch den unverhältnismäßig hohen Aufwand einer kompletten Filmproduktion mit sich zu bringen.

2.1.1 Storyboard

In einem ersten Schritt wurden verschiedene Sequenzen durchgespielt, die einerseits eine einfache Kombination des Schauspielers mit einer künstlich erstellten Umgebung zulassen sollten und andererseits in möglichst kurzer Zeit einen attraktiven stereoskopischen Effekt darstellen konnten. Aufgrund der Zusage eines befreundeten Karate-Sportlers zur Mitarbeit, fiel die Entscheidung dann auf die Darstellung einer Serie von Schlag- und Trittechniken. Ziel sollte es sein, den Zuschauer durch die auf ihn gerichteten Kampftechniken zu beeindrucken.

Der Ablauf der Einstellung und der gewünschte stereoskopische Effekt mussten möglichst exakt festgelegt werden, da hiervon sowohl die Position des Darstellers und dessen Aktionsradius, als auch die Distanz zur Kamera abhing. Um dies im Vorfeld eindeutig zu definieren, wurden mehrere Storyboards gezeichnet. Da es im stereoskopischen Film neben der x- und y-Achse noch eine dritte Achse, nämlich die Tiefe in z-Richtung als Gestaltungsmittel gibt, musste - im Gegensatz zu gewöhnlichen Storyboards - das geplante Scheinfenster eingezeichnet werden. Nur auf diese Weise ließ sich zweifelsfrei erkennen, welchen Tiefeneffekt die geplante Einstellung enthalten sollte. Abbildung 2.1 und 2.2 zeigen einen Auszug aus dem Storyboard der letztendlich ausgewählten Einstellung. Um festlegen zu können, welche Elemente sich im Vorder-, Mittel- oder Hintergrund befinden, wurde das Scheinfenster in die Zeichnungen integriert. In der oberen Zeichnung ist deutlich zu erkennen, dass sich der Darsteller unmittelbar hinter dem Scheinfenster, also im Mittelgrund, befindet. In der unteren Zeichnung hat der Karate-Kämpfer die vor ihm schwebende Steinkugel bereits mit einem Fußtritt getroffen.

31 Autodesk® Maya®, eine Software der Firma Autodesk®, ist eine vielfach ausgezeichnete Lösung für 3D-Modellierung, Animation, visuelle Effekte und Rendering. Weiterführende Informationen zu Maya sind online verfügbar: <http://www.autodesk.de/adsk/servlet/index?siteID=403786&id=9906730>.

32 Microsoft® Excel® ist ein Tabellenkalkulations- und Analyseprogramm. Nähere Produktinformationen sind online verfügbar: <http://office.microsoft.com/de-at/excel/HA101656321031.aspx>.

Wie anhand der relativen Position der entstandenen Bruchstücke zum Scheinfenster leicht zu erkennen ist, fliegen diese auf den Zuschauer zu. Ohne das eingefügte Scheinfenster würden entscheidende Informationen über den geplanten Dreh verloren gehen, was die Kommunikation zwischen Kameramann und Regisseur erschweren würde. Um das Scheinfenster direkt in den Abbildungen platzieren zu können, entstanden die Zeichnungen für das Storyboard direkt in Adobe Photoshop mit Hilfe eines Wacom³³ Grafiktablets.

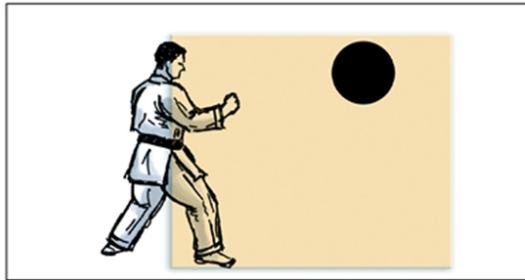


Abbildung. 2.1: Erster Storyboardauszug mit integriertem Scheinfenster



Abbildung. 2.2: Zweiter Storyboardauszug mit integriertem Scheinfenster

2.1.2 Verfügbares Equipment

Bevor die notwendigen stereoskopischen Berechnungen angestellt werden konnten, mußte zunächst das zur Verfügung stehende Kamera-Equipment bekannt sein. Schließlich hängt der gesamte Bildaufbau am Set unter anderem von der minimal erreichbaren Stereobasis der Kameras ab, die aber je nach Kameramodell zum Teil erheblich variiert. Folglich muss als Grundlage der Berechnungen zunächst der zu

³³ Die Firma Wacom Co. Ltd., wurde 1983 in Japan gegründet und hat heute Niederlassungen in den USA und Europa. Wacom entwickelte die ersten Grafiktablets mit druckempfindlichen Stift ohne Batterie und Kabel. Informationen zu den Produkten gibt es online unter: <http://www.wacom-europe.com>.

verwendende Kameratyp mit der dazugehörigen minimalen Stereobasis bekannt sein. Im vorliegenden Fall kamen zwei Sony HVR-Z1 E³⁴ zum Einsatz. Deren minimale Stereobasis liegt bei 15,5 Zentimetern³⁵, die minimale Brennweite f der verwendeten Objektive beträgt 32,5 Millimeter. Mit diesen Daten kann die Scheinfensterweite, also die Entfernung des Scheinfensters von der Kamera, berechnet werden.

2.1.3 Stereoskopische Berechnungen

Auf Grundlage der Storyboards wurden mit dem Darsteller mehrere Einstellungen vereinbart. Um die gewünschten stereoskopischen Effekte zu erhalten, mussten im Vorfeld die jeweilig benötigten Entfernungen des Schauspielers zur Kamera berechnet werden.

Diese Entfernung lässt sich direkt von der Position des Scheinfensters ableiten. Wie in Abbildung 2.1 und 2.2 zu sehen ist, war es geplant, den Darsteller kurz hinter dem Scheinfenster zu platzieren. Also musste zunächst die Scheinfensterweite ermittelt werden.

Da zum Zeitpunkt des Drehs der benötigte Vergrößerungsfaktor (p) noch nicht bekannt war, wurde zur Sicherheit eine maximal tolerierbare parallaxtische Verschiebung (v) von nur einem Millimeter angesetzt. Dadurch berechnet sich die Scheinfensterweite durch die Multiplikation von Stereobasis mit der benutzten Brennweite:

$$\text{Scheinfensterweite (SFW)} = \text{Stereobasis (SB)} \times \text{Brennweite (f)}$$

Im konkreten Fall war die Stereobasis 15,5 cm und die Brennweite 32,5 mm, woraus sich eine Scheinfensterweite von etwa 5,04 m ergibt. Das bedeutet, dass der Darsteller zur Kamera einen kontinuierlichen Mindestabstand von fünf Metern einhalten muss! Bereits an dieser Stelle war somit klar, dass durch den großen Mindestabstand keine Nahaufnahmen möglich sein würden.

2.1.4 Planung der CG-Umgebung

Um eine möglichst problemlose Integration des Darstellers in eine computergenerierte Umgebung zu bewerkstelligen, muss bereits vor den Realaufnahmen die Beschaffenheit und die Lichtstimmung des angestrebten finalen Gesamtbildes bekannt sein. Dementsprechend muss zunächst das Aussehen und die Beleuchtung der CG-Environment festgelegt werden. Wichtige Fragen, die hierbei geklärt werden müssen sind beispielsweise:

- Findet die Einstellung im Freien oder in einem Raum statt (outdoor oder indoor)?

34 Technische Informationen zu dieser Kamera sind im Internet verfügbar: http://www.sony.de/biz/view/ShowProduct.action?product=HVR-Z1E&site=biz_de_DE&pageType=Features&imageType=Main&category=HDVCamcorders.

35 Vgl. Frank, Anja 2007: Bachelor-Thesis Stereoskopische High Definition Produktion, S. 43.

- Herrscht dementsprechend natürliches oder künstliches Licht vor?
- Zu welcher Tages- bzw. Nachtzeit spielt die Einstellung (Schatten)?
- Findet eine Interaktion zwischen Real- und CG-Elementen statt?

Da für ein ausgefeiltes Lichtdesign im Rahmen dieser Arbeit keine Zeit übrig war, wurde festgelegt, dass die Einstellung am frühen Nachmittag im Freien in einem relativ diffusen Licht spielen sollte.

2.2 Realaufnahme

Die Realaufnahmen entstanden im Studio der Fakultät Digitale Medien. Da der Darsteller im finalen Compositing komplett zu sehen sein sollte, musste auch der Boden des Studios mit blauem Stoff ausgelegt werden. Anschließend wurde der Bluescreen möglichst gleichmäßig ausgeleuchtet, um beim späteren Chroma-Keying eine hohe Qualität zu erreichen.

Für den stereoskopischen Kameraaufbau kamen, wie bereits erwähnt, zwei Sony HVR-Z1 E zum Einsatz, die mittels einer Stereoschiene³⁶ nebeneinander auf ein Stativ montiert wurden. Leider sind diese Kameras nicht in der Lage, das in den Vorüberlegungen favorisierte 720p Format aufzuzeichnen. Aus diesem Grund musste im 1080i Format auf MiniDV-Bänder mit MPEG-2-Kompression aufgezeichnet werden, was bereits zu diesem Zeitpunkt auf Probleme beim späteren Chroma-Keying schließen ließ. Um darüber hinaus auch noch die Auswirkungen des Color Spills beim Keying-Verfahren testen zu können, wurde beschlossen, dass der Schauspieler beim Dreh seinen weißen Karateanzug tragen sollte.

Um die richtige Position für die Kameras zu finden, wurde das Set vermessen. Die errechnete Scheinfensterweite lag bei 5,04 Metern. Dazu mussten nach einigen Tests etwa 1,5 Meter Bewegungsspielraum für den Darsteller addiert werden. Daraus resultierte ein Abstand der Kameras zum Bluescreen von mindestens 6,5 Metern, was dazu führte, dass die Kameras am äußersten Rand des Studios aufgestellt werden mussten.

Die Position des Scheinfensters wurde auf dem Boden mittels Klebeband markiert. Anschließend mussten die beiden Kameras absolut synchron eingestellt werden. Dazu wurden alle Automaten deaktiviert und jeder Menüeintrag der Kameras nochmals auf eine identische Einstellung überprüft³⁷.

Bevor mit den eigentlichen Dreharbeiten begonnen werden konnte, war es zwingend erforderlich, alle relevanten Daten festzuhalten, die später in der Postproduktion benö-

36 Diese Stereoschiene ist ein Prototyp, der von Anja Frank im Rahmen einer Bachelor-Thesis konzipiert wurde. Vgl. Frank, Anja 2007: Stereoskopische High Definition Produktion, S. 46-49.

37 Vgl. Frank, Anja 2007: Stereoskopische High Definition Produktion, S. 35.

tigt werden würden. Dazu gehören insbesondere Stereobasis, Scheinfensterweite, Kamerahöhe und Kamerawinkel, sowie alle technischen Kameradaten, wozu in erster Linie Brennweite, Blende und Verschlusszeiten gehören. Außerdem wurden alle Abstände der Schauspieler und Objekte zur Kamera vermessen und notiert. Diese Daten müssen später in der 3D-Software für die virtuelle Kamera exakt übernommen werden. Um dabei keine Angaben zu außer Acht zu lassen, würde sich ein standardisiertes Formular anbieten.

Es folgte eine detaillierte Einweisung des Schauspielers. Dazu gehörte vor allem die klare Festlegung seiner Position und seines erlaubten Aktionsradius anhand von Markierungen auf dem Boden.

Letztendlich wurden neun verschiedene Einstellungen gedreht. Um aus diesen Testaufnahmen das Maximum an Basismaterial und Information zu extrahieren, wurde jede dieser Einstellungen in mindestens drei Varianten wiederholt. Bei der ersten Version befand sich der Darsteller auf der errechneten Position des Scheinfensters. In Version zwei und drei befand er sich jeweils einen halben Meter davor bzw. dahinter. Mit diesem Verfahren konnte in der Postproduktion die Richtigkeit der Berechnungen praktisch überprüft und gegebenenfalls korrigiert werden.

Bei den Aufnahmen der einzelnen Einstellungen wurde immer eine Regieklappe benutzt, anhand derer die rechte und linke Videospur in der Postproduktion synchronisiert werden konnten. Nach dem Dreh fand bereits im Studio eine erste Sichtung des Materials statt. Anschließend wurden die Rohdaten im .m2v-Format³⁸ auf einer mobilen Festplatte gespeichert.

2.3 Postproduktion

In den folgenden Abschnitten werden deshalb anhand praktischer Tests konkrete Erkenntnisse zu Möglichkeiten und Problematiken verdeutlicht.

2.3.1 Synchronisierung der Spuren

Im ersten Schritt der Postproduktion müssen die Bilder der rechten und linken Videospur immer präzise synchronisiert werden, da sonst eine stereoskopische Wiedergabe nicht funktionieren kann.

Die Synchronisierung fand in After Effects CS3³⁹ statt. Es eine neue Komposition erstellt und die .m2v-Dateien der rechten und linken Kamera als Footage importiert.

38 Eine .m2v-Datei ist ein MPEG-2 Elementary Video Stream, also ein direkter Video Encoder-Output.

39 After Effects® gilt als einer der Industriestandards in den Bereichen Motion Graphics und Visual Effects. Nähere Produktinformationen sind online verfügbar: <http://www.adobe.com/de/products/aftereffects/features/> Die Software wurde von der Firma Adobe® entwickelt. Weitere Informationen zum Firmenprofil: <http://www.adobe.com/de/aboutadobe/>.

Um eine bestimmte Einstellung zu extrahieren, wurde das Footage mit In und Out Points versehen und anschließend in der Zeitleiste anhand der in beiden Spuren sichtbaren Regieklappe synchronisiert. Die ausgewählten und synchronisierten Einstellungen wurden im letzten Schritt in maximaler Qualität ohne Komprimierung als .avi-Datei exportiert.

2.3.2 Maskierung

Die Erstellung von qualitativ hochwertigen Masken ist der Schlüssel zu einem überzeugenden Compositing. Dies gilt uneingeschränkt auch für eine stereoskopische Produktion. In diesem ersten Ansatz stand dementsprechend das Ziel im Vordergrund, mittels Chroma Keying und Rotoscoping eine möglichst saubere Maske für den Darsteller zu generieren. Aufgrund der doppelt vorhandenen Videospuren wurde der ohnehin schon aufwändige und zeitintensive Vorgang nochmals massiv erschwert. Die folgenden Schritte und Einstellungen haben sich im Laufe der Testreihen bewährt.

2.3.2.1 Erstellung einer Garbage Matte

Im allerersten Maskierungsschritt sollte eine grobe Maske mit dem Zeichenstift-Werkzeug eingezeichnet werden. Diese sogenannte Garbage Matte⁴⁰ muss angelegt werden, um die Teile des Sets grob von Hand zu maskieren, die nicht vor Green- oder Bluescreen gedreht werden konnten und im finalen Compositing nicht zu sehen sein sollen. Dabei muss aber beachtet werden, dass diese erste grobe Maske für die gesamte Einstellung gilt. Es muss also sichergestellt werden, dass die Maskenumrisse den Darsteller zu keinem Zeitpunkt der Bewegungssequenz beschneiden. Innerhalb dieser groben Maske kann das automatische Chroma-Keying anschließend problemlos durchgeführt werden.



Abbildung. 2.3: Gesamtansicht der originalen Aufnahme

40 Der Begriff Garbage Matte setzt sich aus den engl. Wörtern garbage (dt.: Müll) und matte (dt.: Maske) zusammen und beschreibt einen ersten groben Maskiervorgang von Hand.



Abbildung. 2.4: Erstellung einer Garbage Matte

Abbildung 2.3 stellt die ursprüngliche Gesamtansicht der Aufnahme dar. Abbildung 2.4 zeigt die Aufnahme nach Anwendung einer Garbage Matte. Aufgrund der verwendeten minimalen Brennweite und dem großen Abstand zum Darsteller war eine solche Garbage Matte offensichtlich unumgänglich.

2.3.2.2 Chroma-Keying

Das Chroma-Keying wurde in After Effects mit Hilfe von Keylight 1.2, einem ausgesprochen leistungsfähigen Plug-In der Firma The Foundry, durchgeführt, das auch von vielen professionellen Effekthäusern verwendet wird⁴¹. Da das Erstellen von Masken auch bei einer stereoskopischen Produktion prinzipiell unumgänglich ist, wird darauf in diesem Abschnitt etwas ausführlicher eingegangen.

Um das Chroma-Keying durchzuführen, muss zunächst die gewünschte Videospur ausgewählt werden. Anschließend wird unter Effekt > Keying > Keylight das Keylight Plug-In aufgerufen. Die idealen Einstellungen dieses Effekts hängen selbstverständlich vom verwendeten Footage ab. An dieser Stelle werden deshalb lediglich die grundsätzlichen Möglichkeiten der verfügbaren Effektoptionen angesprochen, wobei sich die im Folgenden verwendete Reihenfolge der Einstellungen in der Praxis bewährt hat:

1. Unter Screen Color muss mit der Pipette im Footage die zu keyende Farbe ausgewählt werden. Dies sollte möglichst nah am Darsteller geschehen.
2. Die Option Screen Pre-blur weicht die Kanten vor dem eigentlichen Keying auf und vermindert damit Kompressionsartefakte. Empfohlener Einstiegswert: 2 bis 3.

41 The Foundry Visionmongers Ltd. ist eine international bekannte Entwicklerfirma für hochspezialisierte Visual Effects Software, mit Sitz in London. Für Keylight, das als Plug-In für viele Softwarepakete (z.B. After Effects, Combustion, Nuke, Avid DS) verfügbar ist, hat The Foundry einen Academy Award erhalten. Kinofilme, an denen mit Keylight gearbeitet wurde, sind beispielsweise: Harry Potter, Die Another Day, Blade 2, Armageddon, Mission Impossible und Last Action Hero. Nähere Informationen zu den Produkten sind unter <http://www.thefoundry.co.uk/index.aspx> verfügbar.

3. Um die entstandene Maske besser beurteilen zu können, sollte jetzt unter view die Option Status ausgewählt werden. In dieser extremen Ansicht werden ausschließlich die Bildteile, die zu 100% deckend bleiben, weiß dargestellt und nur die Bildteile, die zu 100% transparent werden, sind schwarz. Zwischenstufen werden einheitlich in einem mittleren Grau visualisiert, wodurch mangelhafte Maskenteile sofort sichtbar werden. Ziel ist es nun, eine reine schwarz-weiß-Maske zu generieren. Dazu wird zunächst die Option Screen Gain, die in etwa dem Eingangspegel des Hintergrundes entspricht, solange erhöht, bis es um den Darsteller möglichst keine grauen Stellen mehr gibt.

4. Im nächsten Schritt sollten die halbtransparenten Stellen im Darsteller entfernt werden. Dies geschieht mit den Optionen Clip Black und Clip White, die unter dem Menü Screen Matte zu finden sind. Mit diesen beiden Optionen werden die Schwarz- und Weißwerte wie bei einer Tonwertkorrektur beschnitten. Der Wert von Clip White muss nun dementsprechend soweit abgesenkt werden, bis in der Fläche des Darstellers keine Grauwerte mehr vorkommen. Mit Clip Black können die Kanten verstärkt werden.

5. Wie Abbildung 3.7 zeigt, stellt Keylight einen eventuell noch vorhandenen Color Spill, eine bei den Aufnahmen entstandene Farbüberstrahlung des Blue- oder Greenscreens, in der Status-Ansicht mit einer entsprechenden Einfärbung auf dem Darsteller dar. Keylight ist aufgrund einer integrierten Spill Suppression-Funktion grundsätzlich in der Lage, Farbüberstrahlungen selbständig zu korrigieren. Mit den Optionen Replace Method und Replace Color kann aber auch manuell in diesen Farbkorrekturprozess eingegriffen werden. Dazu kann mit der Pipette ein Farbbereich gewählt werden, der explizit von der Software unterdrückt werden soll. In der Regel sind dies typischerweise dunkelgraugrüne oder dunkelgraublaue Bereiche.

6. Für eine Feinabstimmung der Maske sollte ab jetzt unter view auf Final Result umgeschaltet werden. Um die Maske an den geplanten neuen Hintergrund anzupassen, muss dieser nun in die Komposition geladen und in der Spur unter dem maskierten Video platziert werden. Mittels der Option Clip Rollback werden nun eventuell aufgetretene Matte Lines eliminiert. Als Einstieg seien hier Werte von 2 bis 4 empfohlen.

7. Anschließend bietet es sich an, die Maskenränder per Screen Softness etwas weicher zugestalten, was in der Regel eine Integration des Darstellers in das CG-Set unterstützt. Außerdem kann die gesamte Maske unter Screen Shrink/Grow verkleinert bzw. vergrößert werden. Hier bieten sich Werte von -0,5 bis +0,5 Pixel an. Abbildung 2.6 zeigt das finale Chroma-Keying, das offensichtlich noch Schwachstellen aufweist.

Wie in der Detailansicht des Chroma-Keyings in Abbildung 2.7 gut zu sehen ist, ist das finale Chrom-Keying keinesfalls perfekt. Ursache hierfür ist zunächst der Schattenwurf der Stofffalten am Boden, die durch die Trittbewegungen des Darstellers entstanden sind.

Ein weiteres gravierendes Problem war die aufnahmetechnisch bedingte kurze Distanz des Darstellers zum Bluescreen. Der dadurch entstandene tiefe Schatten verunreinigte die Screenfarbe so sehr, dass ein akzeptables Keying-Ergebnis per Software nicht zu erreichen war. Das dritte Problem, welches das automatische Keying massiv erschwerte, war die deutliche Artefaktbildung aufgrund der Interlaced-Aufnahmen in Kombination mit der MPEG-2-Kodierung.



Abbildung. 2.5: Die noch unbearbeitete Status-Ansicht



Abbildung. 2.6: Das finale Chroma-Keying mit Keylight



Abbildung. 2.7: Detailansicht des Chroma-Keyings

2.3.2.3 Rotoscoping

Aufgrund des mangelhaften Chroma-Keyings mussten alle Masken von Hand per Rotoscoping nachgebessert werden. Um dies zu ermöglichen, wurden beide Videospuren als unkomprimierte Einzelbildsequenzen im .tif-Format⁴² mit Alphakanal exportiert (Datei > Exportieren > Bildsequenz). Bei einer Gesamtdauer der Einstellung von 2 Sekunden und einer Bildfrequenz von 25 Bildern pro Sekunde ergaben sich durch die zwei verschiedenen Ansichten insgesamt 100 zu bearbeitende Einzelbilder.

Die Alphakanäle der exportierten Einzelbildsequenzen wurden in Adobe Photoshop im Kontrast verstärkt und im Tonwertumfang reduziert. Anschließend wurde jedes Bild einzeln von Hand per Grafiktablett nachgebessert. Abbildung 2.8 zeigt den Alphakanal eines Einzelbildes vor der manuellen Nachbearbeitung. Das Ergebnis der Nachbearbeitung ist in Abbildung 2.9 dargestellt.

Als größtes Problem beim Rotoscoping erwies sich der Faltenwurf des blauen Stoffes auf dem Studioboden, der durch die Bewegungen des Darstellers entstanden war. Durch die Stofffalten wurden die Füße teilweise verdeckt, wodurch Bildinformationen unwiederbringlich verloren gegangen sind. Damit hat sich das Auslegen von Stoff für BluescreenAufnahmen dieser Art als ausgesprochen unpraktikabel erwiesen.

⁴² Das Tagged Image File Format (TIFF) dient zur Speicherung von Bilddaten, wobei es unterschiedliche Algorithmen zur Datenkompression erlaubt und verschiedene Farbräume nutzen kann. Vgl. TIFF Revision 6.0, 1992. Online verfügbar unter: <http://partners.adobe.com/public/developer/en/tiff/TIFF6.pdf>.



Abbildung. 2.8: Fehlerhafte Maske nach dem Chroma-Keying



Abbildung. 2.9: Korrigierte Maske nach dem Rotoscoping

2.3.3 Erstellung des CG-Sets

Nach Fertigstellung der Masken begann die Erstellung der computergenerierten Umgebung für die Szene in Autodesk Maya⁴³.

Das oberste Ziel bestand darin, das CG-Set unter exakt denselben stereoskopischen Gesichtspunkten zu generieren, unter denen auch die Realaufnahmen gedreht wurden. Um dies zu erreichen, mussten die festgehaltenen Kameradaten des Realdrehs in die virtuellen Kameras der 3D-Software übernommen werden. Somit war es im ersten Schritt notwendig, den realen stereoskopischen Kameraaufbau in Maya nachzustellen. Da es an dieser Stelle immer eine Überschneidung zwischen derameratechnik der realen Welt und der 3D-Computergrafik gibt, werden im Folgenden die entsprechenden Schnittstellen genauer untersucht und verglichen.

⁴³ Maya, eine Software der Firma Autodesk, ist eine vielfach ausgezeichnete Lösung für 3D-Modellierung, Animation, visuelle Effekte und Rendering. Weiterführende Informationen zu Maya sind online verfügbar: <http://www.autodesk.de/adsk/servlet/index?siteID=403786&id=9906730>.

2.3.3.1 Das Rendering

Das Rendering, im deutschen Sprachraum oft Bildsynthese genannt, ist der letzte Schritt im 3D-Erstellungsprozess. Wenn alle Objekte und Lichter der Szene ausgerichtet sind und die gewünschten Kameraeinstellungen festgelegt sind, erfolgt mittels einer Rendering-Software die Erstellung zweidimensionaler Bilder aus dreidimensionalen Szenenbeschreibungen vom Betrachtungspunkt einer im Computer simulierten Kamera aus. Eine Szene zu rendern bedeutet also, aus dem Blickwinkel einer virtuellen Kamera auf die erstellte Szene zu schauen, das sichtbare Blickfeld zu berechnen und das Ergebnis in einer Bilddatei abzuspeichern.

2.3.3.2 Die virtuelle Kamera

Die Überführung einer 3D-Szene in ein zweidimensionales Bild geschieht mit Hilfe mathematischer Projektionsverfahren⁴⁴. Eine virtuelle Kamera in der Computergrafik ist, genau betrachtet, also lediglich eine Repräsentation für bestimmte Projektionsverfahren. Die optischen Eigenschaften einer virtuellen Kamera entsprechen prinzipiell denen einer klassischen Lochkamera, weshalb sie sich in Anlehnung an BENDER & BRILL⁴⁵ durch folgende wesentliche Eigenschaften auszeichnen:

1. Die Kameraposition wird durch die Koordinaten eines Punktes angegeben.
2. Der Bildausschnitt ist immer rechtwinklig
3. Der Schärfebereich ist unendlich groß

Die Projektion bildet Punkte des dreidimensionalen Raumes auf Punkte einer gegebenen Ebene ab, d.h., dreidimensionale Objekte können zweidimensional dargestellt werden. Im vorliegenden Fall kommt ausschließlich die Zentralprojektion zum Einsatz. Die Zentralprojektion, im englischen Sprachraum *perspective projection* genannt, entspricht dem menschlichen Sehen, da weiter entfernt liegende Objekte in einem kleineren Bild resultieren als nahe Objekte. Dieser Effekt wird perspektivische Verzerrung genannt. Da dies dem alltäglichen visuellen Erleben entspricht, wird für eine realistische Szenendarstellung die Zentralprojektion verwendet. Bei dieser schneiden sich alle Projektionsgeraden in einem Aug- oder Zentralpunkt Z, der nicht auf der Projektionsebene liegt. Wie in Abbildung 2.10 zu sehen ist, entsprechen die projizierten Bildpunkte den Schnittpunkten von Projektionsgeraden und Projektionsebene.

44 Eine umfangreiche Darstellung der projektiven Geometrie ist zu finden in: Bär, Gert 2001: Geometrie - Eine Einführung für Ingenieure und Naturwissenschaftler. Eine detaillierte Beschreibung der Parallel- und Zentralprojektion enthält: Bender, Michael & Brill, Manfred 2006: Computergrafik.

45 Vgl. Bender, Michael & Brill, Manfred 2006: Computergrafik, S. 26.

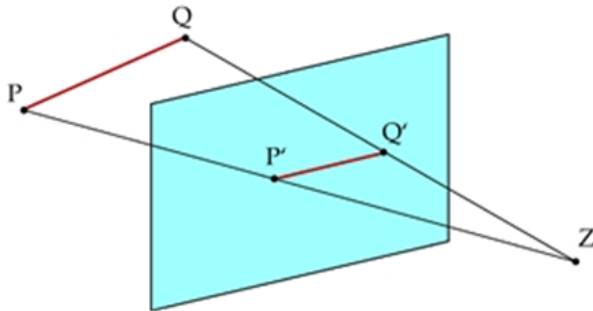


Abbildung. 2.10: Die Zentralprojektion

Zur Definition einer virtuellen Kamera im Raum werden folgende Parameter benötigt⁴⁶:

1. Die Position des Aug- oder Zentralpunktes Z im Weltkoordinatensystem. Z entspricht dabei der Kameraposition im Weltkoordinatensystem.
2. Die Sichtrichtung wird definiert durch den Punkt Z und den sog. look-at-point H .
3. Durch H und Z ist der Vektor n gegeben. n ist die Normale der Projektionsebene. Es ist zu beachten, dass die Sichtrichtung der negativen n -Richtung entspricht.
4. Der Vektor vup (view-up-vector) gibt an, wo im Bild oben ist.
5. Aus vup werden mit Hilfe des Vektorprodukts die normalisierten Vektoren u und v berechnet. u steht dabei für rechts, v steht für oben.

Um dieses Kameramodell im Computer realisieren zu können, muss zwingend der sichtbare Teil der Welt spezifiziert werden, da bisher beispielsweise alle Objekte, ob vor oder hinter der Kamera, abgebildet werden. Dazu wird zunächst ein Rechteck aus den minimalen und maximalen Koordinaten der Vektoren u und v definiert.

Objekte, deren Koordinaten nach der Projektion außerhalb dieses Rechtecks liegen,

⁴⁶ Vgl. Bender, Michael & Brill, Manfred 2006: Computergrafik, S. 36-39.

werden ausgeblendet, bzw. geclippt⁴⁷. Objekte hinter der Kamera werden durch eine Einschränkung der n -Koordinaten ausgeblendet. Um Objekte, die sehr nahe bei der Kamera liegen nicht zu projizieren, werden zwei weitere Elemente eingeführt, die Front- und Batt-Clipping-Plane. Durch diese beiden Ebenen wird ein Intervall $[b, f]$ definiert, was den zulässigen minimalen und maximalen Werten von n -Koordinaten entspricht. Alle Punkte mit n -Werten außerhalb des angegebenen Intervalls werden ebenfalls geclippt⁴⁸.

Durch die angegebenen Spezifikationen wird ein Ausschnitt des dreidimensionalen Raumes definiert. Dieser Raumausschnitt bildet das sogenannte Sichtvolumen oder Viewing Volume, das in Abbildung 2.11 dargestellt ist.

Bei der Zentralprojektion können die Bildgrenzen nicht nur in u und v , sondern auch per Öffnungswinkel angegeben werden. Wie in Abbildung 2.12 zu sehen ist, ist der Öffnungswinkel dabei durch den Winkel zwischen der n -Achse und dem Pyramidenstumpf in v -Richtung definiert.

Man kann den Öffnungswinkel einfach berechnen, sobald die Höhe eines Objekts s im Sichtvolumen und der Abstand d zwischen dem Objekt und dem Zentralpunkt bekannt sind: Öffnungswinkel $\theta = 2 \times \arctan(s/2d)$

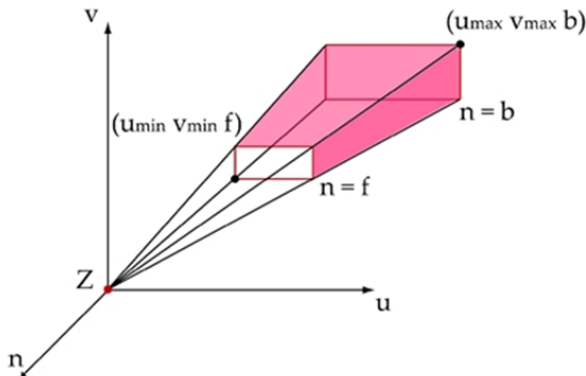


Abbildung. 2.11: Das Sichtvolumen bei der Zentralprojektion

47 Clipping ist ein Verfahren zur Ausschließung oder Beschneidung von Objekten, die ganz oder zum Teil außerhalb des Sichtvolumens liegen. Nur die Teile, die sich innerhalb des Volumens befinden, werden dargestellt. Im Gegensatz dazu kann das Culling nur komplette Objekte von der weiteren Verarbeitung ausschließen. Dies macht dann Sinn, wenn die Objekte bereits vor der Projektion außerhalb des Sichtvolumens liegen. Eine mögliche Objektverdeckung wird von keinem der beiden Verfahren beachtet.

48 Vgl. Bender, Michael & Brill, Manfred 2006: Computergrafik, S. 37.

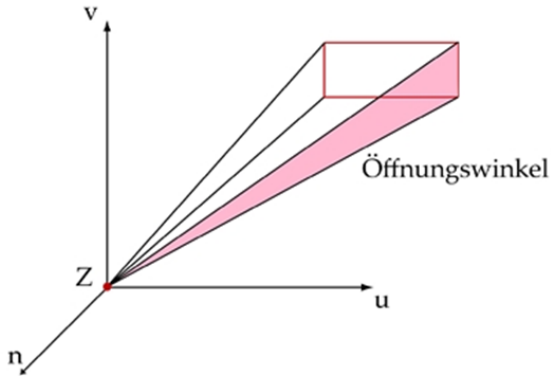


Abbildung. 2.12: Der Öffnungswinkel bei der Zentralprojektion

Der Öffnungswinkel⁴⁹, der auch als Bildwinkel bezeichnet wird, bestimmt, welchen Ausschnitt ein Objektiv von einem bestimmten Standpunkt aus erfassen kann. Der zur Verfügung stehende Bildwinkel hängt von der Brennweite und dem verwendeten Aufnahmeformat ab. Die Brennweite (f) ist, vereinfacht ausgedrückt, die Distanz des Brennpunkts (F) vom optischen Mittelpunkt (M) einer ihm zugeordneten Linse. Diese Betrachtung ist für das weitere Vorgehen absolut ausreichend.

Die Formel zur Berechnung des Bildwinkels lautet: Bildwinkel $\alpha = 2 \times \arctan(d/2f)$ d entspricht dem diagonalen Durchmesser des Aufnahmeformats in Millimetern f entspricht der Brennweite in Millimetern.

Um den Bildwinkel für ein Objektiv mit 50 mm Brennweite zu berechnen, muss zunächst das verwendete Aufnahmeformat bekannt sein. Das Negativ eines KB-Films⁵⁰ hat beispielsweise die Dimension 24×36 mm, woraus eine Diagonale des Aufnahmeformats von 43,27 mm resultiert. Daraus ergibt sich ein Bildwinkel α von $46,8^\circ$.

In den Einstellungsoptionen für virtuelle Kameras sind in der Regel nicht nur der Öffnungswinkel, sondern auch die Brennweite, im englischen Sprachraum als focal length oder focal distance bezeichnet, als einstellbare Option vorhanden. Diese beiden Größen lassen sich verrechnen. Es gilt: Je größer der Öffnungswinkel, desto kleiner die Brennweite. Grundsätzlich erzeugen Objektive mit langer Brennweite bei konstantem Negativformat einen kleineren Bildausschnitt als kurzbrennweitige Objektive. Eine

49 Der Öffnungswinkel wird im engl. Sprachraum als Field of View oder Angel of View bezeichnet.

50 KB-Film steht für Kleinbildfilm. Im Fotografiebereich ist das typische Aufnahmeformat des Kleinbildfilms 24×36 mm. Der 35 mm breite Filmstreifen mit einer doppelseitigen Perforation für den Filmtransport, wurde 1893 von dem schottischen Ingenieur WILLIAM DICKSON (1860-1935) eingeführt und ursprünglich ausschließlich für Kinofilmproduktionen verwendet.

Brennweite wird als normal bezeichnet, wenn sie der Negativformatdiagonale entspricht. Sie gilt als kurz, falls sie höchstens so lang wie die längere Formatseite ist und als lang, wenn sie mindestens doppelt so lang ist wie die längere Formatseite. In der untenstehenden Tabelle 2.1 sind zur Orientierung verschiedene Öffnungswinkel den typischen Brennweiten aus dem KB-Bereich gegenübergestellt.

Öffnungswinkel	Brennweite
84,1°	24 mm
75,4°	28 mm
63,4°	35 mm
46,8°	50 mm
28,6°	85 mm
18,2°	135 mm
12,3°	200 mm

Tabelle 2.1: Öffnungswinkel und Brennweiten

2.3.3.3 Die virtuelle Kamera in Maya

Virtuelle Kameras verhalten sich in Maya wie Kameras der realen Welt und können auch auf ganz ähnliche Art und Weise benutzt werden. Um ein bestmögliches Ergebnis zu erzielen, sollten die Kameraattribute in Maya der verwendeten Kamera des Realdrehs möglichst genau entsprechen. Folgende Attribute müssen zwingend angepasst werden:

- Angle Of View entspricht dem Öffnungswinkel der Kamera. Mit diesem Wert wird in Grad festgelegt, wie groß der sichtbare Bildausschnitt sein soll.
- Mit dem Focal Length-Slider kann die benötigte Brennweite in Millimetern angegeben werden.
- Near Clip Plane und Far Clip Plane beschreiben die minimal bzw. maximal zulässige Entfernung von der Kamera. Objekte, die außerhalb dieser Planes (Ebenen) liegen, werden nicht mehr dargestellt.
- Unter Camera Aperture wird die Höhe und Breite des virtuellen Negativformats in Inches⁵¹ eingestellt. Diese Einstellung ist deshalb entscheidend, weil die Brennweite sich direkt auf dieses angegebene Format bezieht. Der Standardwert von $1,417 \times 0,945$ Inches entspricht dem Kleinbildformat mit 36×24 mm.

51 Ein Inch ist eine im englischen Sprachraum benutzte Maßeinheit und entspricht 0,0254 m.

Um auf die Bedürfnisse der Filmindustrie eingehen zu können, bietet Maya neben diesen Standardparametern auch recht spezielle Einstellmöglichkeiten für die virtuelle Kamera an. So können unter der Option Lens Squeeze Ratio Anpassungen für Filme mit anamorphischen Verzerrungen eingegeben werden.

2.3.3.4 Die Umsetzung

Um in Maya mit Daten aus der realen Welt arbeiten zu können, muss zunächst eine Längeneinheit festgelegt und konsequent beibehalten werden. Im konkreten Testfall wurde für eine Einheit in Maya eine Länge von 10 Zentimetern bestimmt.

Im nächsten Arbeitsschritt musste der reale stereoskopische Kameraaufbau in Maya nachgestellt werden. Dazu wurden zwei virtuelle Kameras generiert und gruppiert, deren Einstellungen bezüglich Brennweite, Blende und Aufnahmeformat exakt den Werten der verwendeten Sony-Kameras entsprachen.

Die im Studio gemessene Stereobasis von 15,5 Zentimetern und die Kamerahöhe von 113 Zentimetern wurden in Maya entsprechend der initial festgelegten Längenverhältnisse mit 1,55 und 11,3 Einheiten umgesetzt. Damit war die Überführung des realen Kameraaufbaus in ein virtuelles stereoskopisches Kamera-Rig⁵² vollzogen.

Anschließend erfolgte die maßstabsgetreue Modellierung einer Szene mit mehreren Gebäuden. Wie in Abbildung 2.13 zu sehen ist, wurden diese zueinander versetzt, um auch die Auswirkung monoskopischer Tiefenreize - wie z.B. der Verdeckung - testen zu können. Aus demselben Grund wurde auch eine in die Tiefe führende Reihe von Steinquadern in die Szene eingefügt.

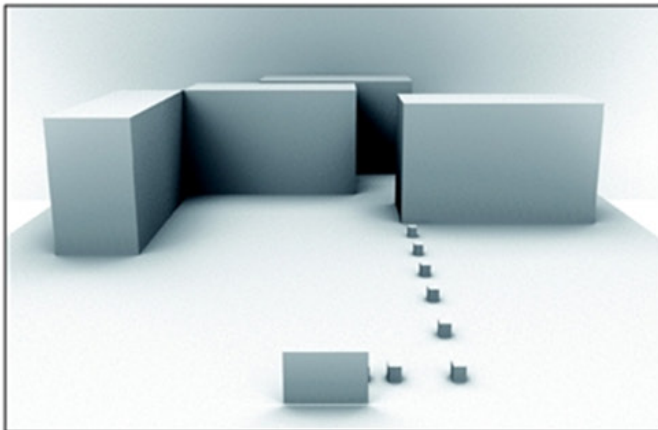


Abbildung. 2.13: Die 3D-Szene ohne Texturen

52 Der englische Begriff Rig bedeutet im deutschen Sprachraum Vorrichtung oder Anlage

Zusätzliche Tiefenreize erhielt die Szene durch die Verwendung einer geeigneten Bodentextur⁵³, die sowohl den monoskopischen Tiefenreiz der perspektivischen Konvergenz, als auch den des Texturgradienten auslöst. In Anlehnung an die im Storyboard verwendete Technik, wurde zur Orientierung das berechnete Scheinfenster aus der realen Welt im 3D-Raum mit Hilfe einer halbtransparenten, orangefarbenen Fläche dargestellt. Abbildung 2.14 veranschaulicht die texturierte und gleichmäßig beleuchtete Szene mit dem eingefügten orangefarbenen Scheinfenster im Ansichtenfenster von Maya. Um jederzeit die Kontrolle über die von den beiden virtuellen Kameras erfassten Bildbereiche zu behalten, wurde ein neues Ansichtenlayout erstellt, das im oberen linken Ansichtsfenster den Bildausschnitt der linken Kamera und im oberen rechten Ansichtsfenster den Bildausschnitt der rechten Kamera darstellt. Das untere Ansichtsfenster dient zur Navigation und Bearbeitung und zeigt dementsprechend eine Komplettübersicht über die gesamte 3D-Szene. Abbildung 2.15 stellt dieses Ansichtenlayout dar.



Abbildung. 2.14: Die beleuchtete 3D-Szene mit Texturen

⁵³ Textur (von lat. *textura*: Gewebe) bezeichnet im Bereich der Computergrafik eine Bilddatei, die auf die Oberfläche einer Geometrie gelegt wird, um deren Aussehen gezielt zu verändern.

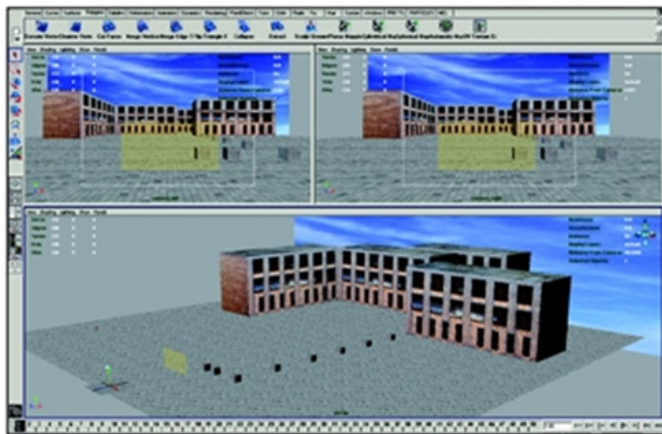


Abbildung. 2.15: Übersicht über die gesamte Szene im Maya

Nach Überprüfung aller Bildinhalte auf korrekte Lage, Perspektive und Parallaxe wurden die Ansichten der linken und rechten Kamera in hoher Qualität mit mental ray⁵⁴ gerendert und separat abgespeichert. Um eine natürlichere Lichtstimmung und einen weichen Schattenwurf zu erhalten, wurde das Rendering mittels Raytracing-verfahren⁵⁵ in Kombination mit Final Gathering⁵⁶ durchgeführt.

Die Animation der zersplitternden Kugel wurde getrennt erstellt. Dieses Vorgehen zählt sich im Compositing aufgrund der größeren Flexibilität bei der Bildkomposition und den schneller durchführbaren Korrekturen der einzelnen Elemente aus.

Nach Festlegung der exakten Position im Bild entstand die Explosion mit Hilfe eines

54 mental ray ist eine Renderingsoftware der Firma mental images GmbH. Die 1986 von Rolf Herken gegründete Firma hat ihren Hauptsitz in Berlin und ist der internationale Marktführer für Rendering- und 3D-Modellierungs-Technologien im Unterhaltungs-, CAD-, Visualisierungs- und Architekturbereich. Weitere Informationen sind online verfügbar: <http://www.mental-images.com/>.

55 Der engl. Begriff Raytracing (dt.: Strahlenverfolgung) bezeichnet einen Renderalgorithmus, der zur Berechnung von photorealistischen Bildern für die Beleuchtung eines Objektpunktes nicht nur das direkt einfallende Licht, sondern auch indirekte Lichtanteile berücksichtigt. Raytracing gehört somit zur Klasse der globalen Beleuchtungsverfahren. Über diese sogenannten Global Illumination-Verfahren gibt es bereits zahlreiche Monographien und Standardliteratur: Sillion, Francois X. & Puech, Claude 1994: Radiosity, Global Illumination; Glassner, Andrew S. 1989: An Introduction to Ray Tracing; Cohen, Michael F. & Wallace, John R. 1993: Radiosity and Realistic Image Synthesis; Fellner, Wolf-Dietrich 1988: Computer-Grafik; Foley, Jim & Dam, Andries van & Feiner, Steven & Hughes, John 1991: Computer Graphics - Principles and Practice. Eine gute Einführung bietet: Bender, Michael & Brill, Manfred 2006: Computer-grafik.

56 Final Gathering ist ein Rendering-Algorithmus, der in Kombination mit dem Raytracing die Qualität von diffusen Schattierungen verbessert.

Partikelsystems. Die Animation wurde anschließend mit den zum Background identischen Rendereinstellungen aus Sicht der linken und rechten Kamera als Sequenz aus jeweils 75 Einzelbildern im .tif-Format mit Alphakanal berechnet und gespeichert. Abbildung 2.16 zeigt ein Bild der finalen Animation aus Sicht der linken Kamera.

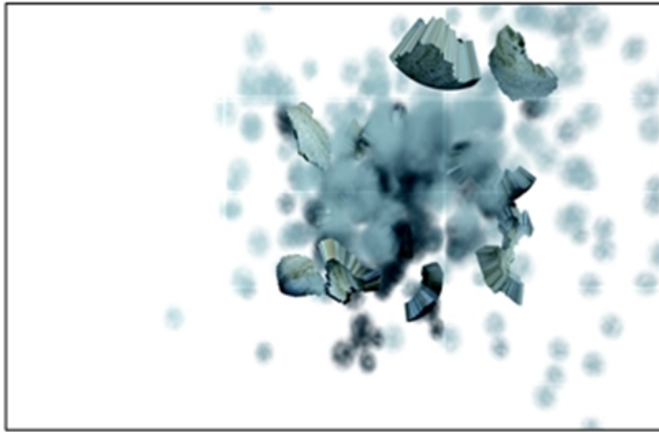


Abbildung. 2.16: Die explodierende Kugel aus Sicht der linken Kamera

2.3.4 Digitales Compositing

Das digitale Compositing wurde mit Adobe After Effects CS3 durchgeführt. Es umfaßte im Wesentlichen das Zusammenfügen und die Feinabstimmung der bisher erstellten 2D-Bildelemente zu einem möglichst homogenen Gesamtbild. Dazu wurden die Bildelemente in separaten Ebenen übereinander geschichtet. Die im Vorfeld maskierten Bildteile sind transparent, so dass Bildteile auf tiefer liegenden Ebenen sichtbar wurden.

Grundsätzlich betrachtet sind die beiden obersten Ziele beim Compositing, den Zuschauer nicht erkennen zu lassen, dass das betrachtete Composite aus mehreren Einzelementen besteht und um die Illusion zu schaffen, dass alle Objekte in der Szene von denselben Lichtquellen beleuchtet wurden⁵⁷. Aufgrund des engen Zeitplanes und der begrenzten Rechenleistung des verwendeten Systems, mussten diese Ziele zugunsten der praktikablen Durchführbarkeit etwas in den Hintergrund treten. Der Fokus lag bei dieser Testproduktion definitiv auf der Ergründung der prinzipiellen Machbarkeit und nicht auf der maximal erreichbaren Compositingqualität. Folgende Arbeitsschritte wurden durchlaufen:

1. Erstellen einer neuen Komposition in After Effects.

⁵⁷ Vgl. Allen, Damian 2007: Encyclopedia of visual effects, S. 6.

2. Einfügen des gerenderten 2D-Himmels der linken Ansicht auf der untersten Ebene.

Um der Szene mehr Lebendigkeit zu verleihen, wurde die Position des Himmelsbildes animiert.

3. Hinzufügen der Gebäude-Ansicht, die mit der linken virtuellen Kamera gerendert wurde. Die Himmelsebene bildet zusammen mit den Gebäuden im fertigen Film den Hintergrund.

4. Platzierung der mit der linken Kamera gedrehten Ansicht des Darstellers in der folgenden Ebene. Dabei ist unbedingt zu beachten, dass die Position des Darstellers im Composite von der Größe des Versatzes zwischen den real gedrehten linken und rechten Ansichten abhängt. Entsprechend dem Parallaxenversatz muss der Darsteller entweder weiter vorne oder weiter im Hintergrund platziert werden. An dieser Stelle sei bereits erwähnt, dass die Anpassung der Parallaxen der verschiedenen Ebenen ohne Zweifel die Hauptdomäne der stereoskopischen Postproduktion darstellt.

5. Einfügen von künstlichem Schatten. Dieser Arbeitsschritt ist ausgesprochen wichtig, da der Schatten wesentlich zur glaubwürdigen Integration des Darstellers in das CG-Set beiträgt. Darüber hinaus stellen Schatten einen monoskopischen Tiefenreiz dar.

6. Im nächsten Arbeitsschritt erfolgt die Anpassung von Vorder- und Hintergrund. Dazu muss in der Regel zunächst die Maskenkante des Darstellers im Vordergrund verkleinert werden, um eventuell vorhandene Artefakte⁵⁸ zu eliminieren. Dies geschieht mit Hilfe der Funktion Maske vergrößern/verkleinern (Effekte > Matte > Matte vergrößern/ verkleinern). Ein Faktor von +0,75 bis +1,5 hat sich während der Testreihen bewährt. Anschließend muss noch eine Tonwertkorrektur durchgeführt werden, um die Schatten von Vorder- und Hintergrund anzugleichen (Effekt > Farbkorrektur > Tonwertkorrektur).

7. Platzierung der Explosionsanimation der linken Ansicht entsprechend den Angaben im Storyboard in der obersten Ebene.

8. Abspeichern der fünf Ebenen als Unterkomposition mit Namen links.

Dieselben Arbeitsschritte müssen auch für die rechten Ansichten durchlaufen werden. Dabei ist unbedingt darauf zu achten, dass Manipulationen oder Animationen in beiden Ansichten absolut synchron durchgeführt werden. Wird diese Regel vernachlässigt, kann aufgrund der Abweichungen keine stereoskopische Fusion entstehen. Der

58 Artefakte an der Maskenkante treten vor allem dann auf, wenn das Originalmaterial im Zeilensprungsverfahren (interlaced) aufgenommen wurde.

Zuschauer wird somit im besten Fall keine räumliche Tiefe wahrnehmen können - im schlimmsten Fall wird er Kopfschmerzen und Übelkeit verspüren. Zuletzt wurde noch eine passende Tonspur hinzugefügt. Diese schafft einerseits Atmosphäre und verstärkt andererseits die Bildaussage und trägt somit in bedeutendem Maße zur Akzeptanz des Composite bei.



Abbildung. 2.17: Ein Einzelbild der fertigen Animation aus der linken Ansicht

Da die beiden fertigen Ansichten im unkomprimierten Zustand bei einer Animationsdauer von 20 Sekunden ein Datenvolumen von insgesamt 4,5 Gigabyte aufwiesen, wurden sie als linke und rechte .avi-Datei im DVCPRO-Format exportiert. Das dadurch entstandene Datenaufkommen von jeweils 35 Megabyte ermöglichte die Echtzeitwiedergabe der Files im Stereoscopic Player. In Abbildung 2.17 ist ein komplettes Einzelbild aus der linken Ansicht der finalen Animation dargestellt.

2.4 Stereoskopische Wiedergabe

Die Überprüfung der stereoskopischen Konsistenz erfolgte während der Testproduktion auf zwei unterschiedlichen Wegen: Direkt in After Effects und mit dem Stereoscopic Player.

2.4.1 Stereoskopische Vorschau in After Effects

Wie die Abbildung 2.18 zeigt, kann in After Effects mit dem Effekt 3D-Brille ein Anaglyphenbild aus zwei ausgewählten stereoskopischen Videospuren direkt im Kompositionsfenster der Software dargestellt werden. Dies ist eine ausgesprochen schnelle und effektive Methode um die stereoskopische Komposition zu überprüfen. Der Effekt ist unter Effekt > Perspektive > 3D-Brille zu finden, er funktioniert allerdings nur im 8-Bit-Farbmodus. Um Darstellungsfehler zu vermeiden, dürfen ausschließlich Ausgangsbilder mit denselben vertikalen Dimensionen verwendet werden. Außerdem

sollten für die Ebeneneigenschaft Position nur ganzzahlige Werte eingegeben werden.

Die Option 3D-Ansicht gibt an wie die linke und rechte Ansicht kombiniert werden. Während der Tests hat sich das Verfahren Symmetrisch farbig rot/blau bewährt, da es nicht wie die Option Rot Blau LR nur die Luminanzwerte der beiden Ebenen zur Verechnung nutzt, sondern die Ebenen unter Verwendung der ursprünglichen RGB-Kanäle in eine stereoskopische Ansicht konvertiert. Dabei bleiben die Originalfarben der Ebenen weitgehend erhalten.

Die Option Konvergenzversatz bestimmt den Betrag, um den die beiden Ansichten gegeneinander verschoben werden. Hier haben sich, in Abhängigkeit vom Ausgangsmaterial und gewünschter Tiefe, Werte zwischen -18 bis -15 bewährt.

Unter Symmetrie können die Luminanzwerte der Farben korrigiert werden, um Artefakte und Schattenbildung zu reduzieren. Praktikable Werte liegen hier zwischen 7 und 12. Je höher der Wert, desto stärker ist das Ausgabeergebnis gesättigt.

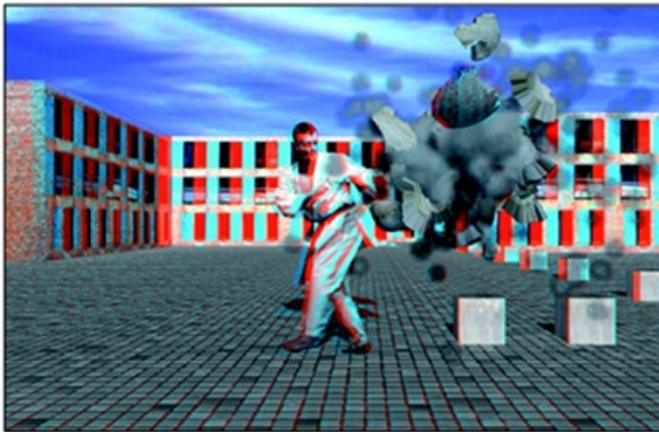


Abbildung. 2.18: Der 3D-Brillen-Effekt in After Effects

2.4.2 Stereoskopische Vorschau mit dem Stereoscopic Player

Neben der direkten Vorschau im Compositing mussten immer wieder einzelne Testsequenzen gerendert werden, um sie in den verschiedenen zur Verfügung stehenden Wiedergabegeräten testen zu können. Zur Wiedergabe dieser generierten Videofiles wurde der Stereoscopic Player benutzt. Dabei handelt es sich um eine auf DirectShow basierende, universell einsetzbare Software zur Wiedergabe von stereoskopischen Videos. Der Player, der zum Zeitpunkt dieser Bearbeitung in der Version 1.3 vorliegt,

wurde von Peter Wimmer, Inhaber der Firma 3dtv.at⁵⁹, entwickelt. Der entscheidende Vorteil dieser Software besteht darin, dass es nicht mehr notwendig ist, für jedes benötigte Wiedergabeverfahren eine entsprechende Videodatei zu erstellen. Stattdessen muss nur noch ein Video erzeugt werden, das der Player dann in Echtzeit in das gewünschte Ausgangsformat konvertiert.

Als Eingangsformat werden dabei vor allem AVI, MPEG, WMV und ASF unterstützt, während als Wiedergabeformat alle gebräuchlichen stereoskopischen Formate, wie z.B. das Anaglyphenformat zur Betrachtung mit Rot-Cyan-Brillen, das Interlaced-Format zur Betrachtung mit Shutterbrillen oder das Nebeneinanderformat zur Betrachtung in Schieltechnik, angeboten werden. Darüber hinaus werden auch autostereoskopische Displays der Firma TRIDELITY direkt unterstützt.

Der Stereoscopic Player setzt zur stereoskopischen Darstellung allerdings Dateien voraus, die eine linke und rechte Ansicht enthalten. Gewöhnliche zweidimensionale Filme können nicht dreidimensional wiedergegeben werden. „Der derzeitige Stand der Technik erlaubt keine Konvertierung in Echtzeit in brauchbarer Qualität. Computer sind nicht in der Lage, den Bildinhalt zu interpretieren und den Vordergrund zuverlässig vom Hintergrund zu trennen.“⁶⁰

Die gerenderten Testsequenzen wurden zunächst per Anaglyphenverfahren und entsprechender Brille betrachtet. Im weiteren Verlauf der Tests wurden mir von der Firma TRIDELITY DISPLAY SOLUTIONS GMBH⁶¹ freundlicherweise zwei autostereoskopische Displays zur Verfügung gestellt. Diese konnten parallel zu einem gewöhnlichen LCD-Display betrieben werden und gestatteten so die Überprüfung der erzielten Ergebnisse in Echtzeit.

2.5 Praktische Erkenntnisse und Lösungen

Die Umsetzung des ersten getesteten Fusionsverfahrens im 2D-Raum brachte viele praktische Erkenntnisse mit sich, die - was den Realdreh und das Keying betrifft - auch für alle weiteren Verfahren von Bedeutung sein werden. Aus diesem Grund sind diese Erkenntnisse bezüglich aufgetretener Probleme und den entsprechenden Lösungsansätzen im Folgenden nochmals detailliert zusammengefasst.

59 Die Firma 3dtv.at hat ihren Sitz in der Wankmüllerhofstr. 9, 4020 Linz, Österreich. Email: office@3dtv.at, URL: <http://www.3dtv.at>.

60 Wimmer, Peter 2008: Homepage von 3dtv.at. URL: <http://www.3dtv.at/Products/Player/>.

61 Die Firma TRIDELITY Display Solutions GmbH mit Sitz in St. Georgen / Schwarzwald hat sich auf die Neu- und Weiterentwicklung von serienreifen autostereoskopischen 3D-Displays spezialisiert. Weitere Informationen sind auf der Firmenhomepage verfügbar: <http://www.tridility.de>.

2.5.1 Storyboard

Das gestalterische Element der Tiefe muss bereits im Storyboard berücksichtigt werden, da es sonst während der Entwicklungsphase der Geschichte nicht genügend beachtet werden kann. Aus diesem Grund ist es zwingend erforderlich, dass das Scheinfenster in das Storyboard integriert wird. Ohne das eingefügte Scheinfenster gehen entscheidende Informationen über den geplanten Dreh verloren, was die Kommunikation innerhalb der Produktionskette (Kunde, Produzent, Regisseur, Cutter und Postproduktion) erschwert.

ROBERT NEUMAN, Stereo Lead der Walt Disney Animation Studios geht bei der Erstellung der Stereo-3D-Storyboards sogar noch einen Schritt weiter. Er vergibt für jede Einstellung einen Stereotiefenwert von minimal 1 für eine geringe Tiefe bis maximal 10 für eine sehr große Tiefe. Der Wert der Tiefe hängt dabei von der emotionalen Tiefe der Einstellung ab. Ist der Zuschauer in einer bestimmten Einstellung mit dem Darsteller emotional verbunden - beispielsweise weil der Darsteller traurig ist und der Zuschauer ihn deshalb am liebsten umarmen möchte - legt NEUMANN einen hohen Tiefenwert fest und läßt den Darsteller damit off-screen erscheinen.⁶²

2.5.2 Lichtstimmung

Der Look und der Aufbau der finalen computergenerierten Umgebung müssen vor dem eigentlichen Realdreh eindeutig festgelegt werden, da die Darstellerbeleuchtung vor dem Greenscreen entsprechend der Lichtstimmung der späteren computergenerierten Umgebung (sonnig, bewölkt, innen, außen, Tag oder Nacht) angepasst werden muss. Dieses Vorgehen erspart viel Zeit bei der Postproduktion und ist Voraussetzung für ein stimmiges Compositing, da die Anpassung des Real-Footage an das CGI deutlich vereinfacht wird.

Darüber hinaus kann der Aufbau der späteren Umgebung bei den Greenscreen-Aufnahmen grob mit grünen Elementen nachgebaut werden, um einen korrekten Schattenwurf des Darstellers zu erzeugen. Dieser Schatten kann beim Keying-Vorgang erhalten bleiben und kann somit im anschließenden Compositing integriert werden. Dadurch erhöht sich der Realitätsgrad der Fusion beträchtlich.

2.5.3 Equipmentvorbereitung

Vor dem Realdreh muss frühzeitig der zur Verfügung stehende Kamerateyp und der mögliche Stereoaufbau der Kameras bekannt sein, da sich daraus der mögliche Brennweitenbereich und die minimale Stereobasis ergeben, von denen alle weitere Berechnungswerte abhängen. Um die Brennweite exakt festlegen zu können, bietet sich der Einsatz von Festbrennweiten an, da bei Zoomobjektiven eine einheitliche Brennwei-

62 M. EMPEY, CG Supervisor und R. NEUMAN, Stereo Lead, Walt Disney Animation Studio: The Use of Stereoscopic Depth as a Storytelling Tool. Vortrag im Rahmen der fmx/08: 13th International Conference on Animation, Effects, Games and Digital Media. May 05-08, 2008 in Stuttgart. URL: www.fmx.de.

tenwahl nur beim maximalen und minimalen Brennweitenwert in den Grenzen der Herstellertoleranz möglich ist. Eine simultane manuelle Einstellung aller weiteren Kameraparameter ist Pflicht - alle automatischen Einstellmöglichkeiten sind zu deaktivieren. Dies gilt im Besonderen für Shutter, Blende, Weißabgleich und Gain.

2.5.4 Das ideale Aufnahmeformat

Das ideale Aufnahmeformat ist bei der gegebenen Aufgabenstellung von den hohen Anforderungen des Chroma-Keying-Verfahrens an das Ausgangsmaterial abhängig. In Kapitel 1.2 wurde bereits festgestellt, dass die erzielbare Keyingqualität direkt von der Signalaufösung der Farbkanäle, der Rauschanfälligkeit der Kanäle und der verwendeten Codierung des Speichermediums abhängt. Die ideale Aufnahmelösung bestünde also darin, die Signale direkt vom Kamerakopf in 10 Bit RGB 4:4:4 unkomprimiert auf Festplatten aufzuzeichnen.

Leider standen für den Realdreh nur zwei Sony HVR-Z1 E Kameras zur Verfügung, die lediglich im 1080i Format auf MiniDV-Bänder mit MPEG-2-Kompression aufzeichnen können, was zu massiven Problemen beim späteren Chroma-Keying führte.

2.5.4.1 Interlaced oder Progressive?

Die Frage, ob im Zeilensprungverfahren (engl.: interlaced scanning) oder im Vollbildverfahren gedreht werden sollte, ist aufgrund der Artefaktbildung, die bei der Interlaced-Aufnahme in Kombination mit der MPEG-2-Kodierung entsteht, schnell zu beantworten. Mit Nachdruck muss hier nochmals darauf hingewiesen werden, dass ausschließlich das Vollbildverfahren für ein qualitativ hochwertiges Keying und Compositing geeignet ist.

2.5.4.2 Anforderungen und Lösungsansätze

Aus den bisherigen Überlegungen resultieren zwangsläufig folgende Anforderungen an ein Aufnahmeformat zur qualitativ hochwertigen stereoskopischen Bluescreen-Aufzeichnung:

- Simultane Aufzeichnung zweier unkomprimierter Videoströme.
- Unkomprimierte Signalverarbeitung im Bildsensor der Kamera
- Aufzeichnung ausschließlich im unkomprimierten Vollbildmodus (progressive), also keinesfalls tapebasiert.
- Aufzeichnung idealerweise mit 10 Bit, um einen Verlust von Farbtiefe und damit eine Reduktion der Farbvarianten pro Farbkanal zu vermeiden.

Anhand dieser Anforderungen ist sofort zu erkennen, dass die Sony HVR-Z1 E Kameras für den durchgeführten Dreh vollkommen ungeeignet waren, da sie kein progressives HDV aufzeichnen können und die Videodaten ausschließlich MPEG-2-komprimiert auf Band gespeichert werden. Die folgende Abbildung 2.19 verdeutlicht das hierbei entstehende Qualitätsproblem anhand einer Detailansicht der Aufnahme. Es

ist leicht zu erkennen, dass an dem schwarzen Karategürtel deutliche quergerichtete Artefakte auftreten, die das Chroma-Keying erschweren.



Abbildung. 2.19: Qualitätsverlust durch das Zeilensprungverfahren

2.5.5 Formular zur Berechnung der Tiefenwerte

Um den Darsteller gemäß den Vorgaben im Storyboard positionieren zu können, müssen alle stereoskopischen Tiefenwerte berechnet und als Anhaltspunkte für Darsteller und Kameramann im Set markiert werden. Da diese Berechnungen bei jedem Dreh mit denselben Formeln durchgeführt werden müssen, entstand die Idee, diesen Vorgang mit einem Excelldokument zu standardisieren. Ziel war es, ein Standardformular für die stereoskopische Produktion zu entwickeln, mit dem einerseits alle notwendigen Berechnungen möglich sein sollten und andererseits alle relevanten Daten des Realdrehs festgehalten werden konnten, die in der Postproduktion benötigt werden. Zu den Daten, die zwingend notiert werden müssen gehören Stereobasis, Kamerahöhe, Kamerawinkel und alle Abstände der Schauspieler und Objekte zur Kamera.

Darüber hinaus müssen selbstverständlich alle technischen Kameradaten erfasst werden. Dazu gehören in erster Linie Brennweite, Blende und Verschlusszeiten. Diese Daten müssen später in der 3D-Software exakt für die virtuelle Kamera übernommen werden, damit das computergenerierte Szenario an die Bedingungen des Realdrehs angepasst werden kann. Steht ein standardisiertes Formular für die Kameradaten am Set zur Verfügung, gehen die Daten nicht verloren.

Folgende Daten müssen in der aktuellen Version des Formulars eingegeben werden:

- Angaben zu den Beteiligten: Projektname, Regisseur, Kameramann, Stereooperator.

- Angaben zum Drehtermin: Datum, Uhrzeit.
- Angaben zur Orientierung im Storyboard: Seite, Szene, Einstellung.
- Kameradaten: Kameramodell, Brennweite, Blende, Kamerahöhe, Kamerawinkel.

Anschließend werden die erforderlichen Berechnungen durchgeführt. Dazu muss zunächst die Scheinfensterweite (SFW) ermittelt werden. Da zum Zeitpunkt des Drehs der benötigte Vergrößerungsfaktor (p) noch nicht bekannt war, wurde zur Sicherheit eine maximal tolerierbare parallaktische Verschiebung (v) von nur einem Millimeter angesetzt. Dadurch berechnet sich die Scheinfensterweite durch die Multiplikation von Stereobasis mit der benutzten Brennweite. Es müssen also lediglich Stereobasis und Objektivbrennweite eingegeben werden, worauf im Feld Scheinfensterweite automatisch der richtige Wert erscheint. Im folgenden Schritt wird die Nahpunktweite (NPW) festgelegt oder ausgemessen. Nach Eingabe des entsprechenden Wertes in das Formular, erscheint die dazu passende Fernpunktweite (FPW). Die Tiefenbereichsweite (TBW) errechnet sich dann durch die Subtraktion von FPW und NPW.

Die hinterlegten Formeln können selbstverständlich der Aufnahmesituation angepasst werden. Wird bei einer Innenaufnahme die Fernpunktweite (FPW) beispielsweise durch das Bühnenbild oder die Studiowand vorgegeben, kann die dazu passende Nahpunktweite (NPW) durch folgende Formel berechnet werden: $NPW (cm) = (SFW (cm) \times FPW (cm)) / (SFW (cm) + FPW (cm))$.

Die untere Hälfte des Formulars ist einer Skizze des Drehortes vorbehalten. Hier können nochmals vor Ort getroffene Änderungen oder Besonderheiten vermerkt werden. Außerdem sollten an dieser Stelle alle gemessenen Distanzen von der Kamera zu den Objekten der Szene eingetragen werden. Diese Daten sind bei der späteren Erstellung der 3D-Umgebung ausgesprochen hilfreich.

2.5.6 Chroma-Keying

Nach bisherigem Kenntnisstand findet der Realdreh immer vor Blue- bzw. Greenscreen statt, weshalb alle Besonderheiten dieser speziellen Aufnahmetechnik zu beachten sind.

- Hochwertiges Aufnahmeformat benutzen: Ausschließlich unkomprimierte Progressiv-Aufnahmen (z.B. 720p) mit maximaler Farbtiefe verwenden. Optimierung der Belichtung und Farbsättigung im Kameramenu unter Waveform/Vektorscope-Kontrolle. Verwendung einer weiten Blendenöffnung, um den Darsteller vom Hintergrund zu lösen.
- Richtige Keyingfarbe auswählen: Grundsätzlich müssen Farbe, Textur und Reflexionsverhalten des zu keyenden Objekts bei der Wahl der Keying-Farbe berücksichtigt werden. Es gilt: besser das Objekt ändern (z.B. Textilien tauschen) und die Keyfarbe dem Aufnahmemedium anpassen.

- Auf eine gleichmäßige Beleuchtung achten: Der Screen muss gleichmäßig ausgeleuchtet werden, damit das automatische Chroma-Keying ein gutes Ergebnis liefern kann. Eine oft formulierte Anforderung ist, „dass nicht mehr als 1/3 Blendenunterschied auf dem Screen messbar ist zwischen der hellsten und der dunkelsten Stelle.“⁶³ Wer schon einmal versucht hat, diese Anforderung auf großen Flächen zu erfüllen weiß, wie schwierig das sein kann. Allgemein gilt: Reflexionen gering halten: weniger direktes Licht auf den Screen, dafür mehr Spitzlichter von hinten, gleichmäßigere Ausleuchtung des Screens, größerer Abstand des Schauspielers zum Screen und weiches Grundlicht.

- Auf die richtige Lichtstimmung achten: Beim Dreh von Einstellungen vor Green- bzw. Bluescreen muss der Beleuchtung besondere Aufmerksamkeit geschenkt werden. Das Licht, welches auf den Vordergrund fällt, sollte im Idealfall mit der Lichtstimmung des späteren Hintergrundes kongruent, d.h. übereinstimmend sein. Dies gilt insbesondere für Lichtrichtung, Lichtintensität und Farbtemperatur. „Das Kriterium [für die gewünschte Übereinstimmung] ist jedoch nicht der Messwert eines Belichtungsmessers, sondern ob es echt aussieht“⁶⁴. Auch aus diesem Grund bietet sich ein HDV⁶⁵-Aufnahmesystem an, das unkomprimiert direkt auf ein Computersystem mit Festplatten-RAID aufzeichnet, da in diesem Fall - genügend Rechenleistung und Arbeitsspeicher vorausgesetzt - der resultierende Effekt simultan begutachtet und ggf. korrigiert werden kann.

- Vermeidung von Color Spill bei der Aufnahme: Wie bereits erwähnt ist Color Spill eine Farbüberstrahlung eines Blue- oder Greenscreens auf helle Objekte im Vordergrund. Dieses Phänomen tritt - wie bei der konkreten Testproduktion geschehen - vor allem bei weißer Garderobe des Schauspielers auf. Da in diesem Fall auch der Vordergrund die Screenfarbe annimmt, wird der Keying-Prozess deutlich erschwert - oder sogar unmöglich gemacht. Obwohl heute bei allen professionellen Keying-Systemen Features zur Eliminierung von Color Spill implementiert sind, gibt es nach wie vor Bildmaterial, das auch mit Hilfe von Software nicht mehr gerettet werden kann. Dies gilt insbesondere für filigrane Strukturen wie Haar und atmosphärische Effekte (Regen, Nebel, Staub, Explosionen...)⁶⁶. Aus diesem Grund muss Color Spill bereits bei der Aufnahme vermieden werden. Dies kann beispielweise dadurch geschehen, dass beim Darsteller auf entsprechend unproblematische Kleidung geachtet wird und er

63 Bertram, Sacha 2005: VFX, S.74.

64 Bertram, Sacha 2005: VFX, S.23.

65 HDV ist ein Akronym für High Definition Video. Dabei handelt es sich um ein in Zusammenarbeit von Canon Inc., Sharp Corporation, Sony Corporation und der Victor Company of Japan, Limited (JVC) entwickeltes Videoformat, das in zwei verschiedenen Aufnahmespezifikationen (HDV 720p und HDV 1080i) auf DV-/Mini DV Band aufgezeichnet werden kann. Weiterführende Informationen sind auf der HDV Information Web Site unter <http://www.hdv-info.org/> zu finden.

66 Vgl. Bertram, Sacha 2005: VFX, S.85.

möglichst weit vom Screen entfernt positioniert wird. Leider war der Bewegungsspielraum im Studio extrem eingeschränkt. Ein begrüßenswerter Abstand zwischen Darsteller und Bluescreen war aufgrund der geringen Studiobreite nicht zu realisieren. Verbleibende Farbüberstrahlungen, z.B. auf den Gesichtern der Darsteller, mussten später in der Postproduktion per Color Grading korrigiert werden.

- Garbage-Matte nutzen: Ist die Screenfläche zu klein, um das gesamte Set zu erfassen, muss nach der Aufnahme manuell eine sog. Garbage Matte (garbage, dt.: Müll) angelegt werden, um die Teile des Sets grob zu maskieren, die nicht vor Grün gedreht werden konnten und im finalen Compositing nicht zu sehen sein sollen. Innerhalb dieser groben Maske kann dann das Chroma-Keying problemlos durchgeführt werden⁶⁷.

2.5.7 Prinzipielle Einschränkungen

Während das menschliche visuelle Wahrnehmungssystem Standbilder mit der Zeit problemlos aufnehmen und fusionieren kann, bekommt das Gehirn bei Bewegungsbildern oft zu wenig Zeit, um sich auf die Illusion einzulassen und sie zu akzeptieren.

In Filmbildern kommt es bei schnellen Bewegungen und schnellen Schnitten demnach zu einer Verminderung des 3D-Effekts. Daraus folgt, dass grundsätzlich weniger Schnitte als bei einer konventionellen Produktion eingeplant werden sollten und bereits beim Schnitt darauf geachtet werden muss, dass die Einstellungslängen den 3D-Effekt beim Betrachter überhaupt entstehen lassen. Aus diesem Grund sollten Einstellungen auch sehr langsam beginnen und erst nach einiger Zeit extreme 3D-Effekte auftreten⁶⁸. Diese Erkenntnis wurde beim finalen Schnitt der Testproduktion umgesetzt. Durch den mehrsekündigen initialen Abschnitt der Animation, bei der sich ausschließlich der Himmel leicht bewegt, bekommt das Gehirn die Gelegenheit sich mit der räumlichen Tiefe vertraut zu machen. Erst wenn dies geschehen ist, folgt der extreme 3D-Effekt der explodierenden Kugel.

2.5.8 Stereo-3D-PlanungsTool

Die anfänglichen Überlegungen in der Planungsphase des ersten Fusionsverfahrens führten bald zu der Erkenntnis, dass zur Umsetzung einer vorab geplanten räumlichen Tiefe in der 3D-Umgebung ein Hilfsmittel notwendig sein würde. Dieses Hilfsmittel sollte mir die Stereokameras automatisch erstellen, die Orientierung im 3D-Raum erleichtern und die Kontrolle über die stereoskopische Tiefenwirkung jedes einzelnen Objektes der computergenerierten Szene sichern. Nach intensiver Recherche ergab sich, dass die Firma VisuMotion⁶⁹ in Jena, Deutschland ein 3D Stereo Camera Plug-in

67 Vgl. Allen, Damian 2007: Encyclopedia of visual effects S. 29; Bertram, Sacha 2005: VFX, S. 74

68 Vgl. Allen, Damian 2007: Encyclopedia of visual effects, S. 496.

69 VisuMotion GmbH, 3D Software und Hardware Solutions. Prüssingstraße 35, 07745 Jena, Deutschland. URL: <http://www.visumotion.com/>.

für Maya anbietet, das dieser Aufgabe gerecht worden wäre.

Leider stellte sich heraus, dass dieses Plug-In im Rahmen der vorliegenden Arbeit zu teuer war, worauf ich mich entschloss ein eigenes Tool zu schreiben. Dieses Tool musste im Ansichtenfenster von Maya folgende Anforderungen erfüllen:

- Automatische Generierung eines Stereokamera-Rigs, bestehend aus zwei gruppierten Kameras, die alle Bewegungen simultan ausführen und identische Einstellungsparameter (Brennweite, Blende, Filmformat...) aufweisen.
- Kontinuierliche numerische Anzeige der verwendeten Stereobasis.
- Visuelle Darstellung der Nahpunktweite, Scheinfensterweite und Fernpunktweite.
- Kontinuierliche numerische Anzeige aller Distanzen zwischen Kameras und den genannten Tiefenweiten und zwischen den Tiefenweiten untereinander.

Die praktische Umsetzung des Tools geschah mit MEL™, der Maya Embedded Language. Dabei handelt es sich um eine direkt in Maya integrierte Scriptsprache, die den Zugriff auf alle Elemente der Software erlaubt. Das erstellte Script wurde anschließend als neuer Button in der Maya-GUI hinterlegt.

Abbildung 2.20 zeigt einen Screenshot des Ansichtenfensters von Maya nach Aktivierung des Stereo-3D-Tools, das alle meine definierten Anforderungen erfüllt. Die beiden Stereokameras wurden automatisch erstellt und gruppiert. Eine numerische Anzeige zwischen den Kameras zeigt kontinuierlich die aktuell verwendete Stereobasis in Maya-Einheiten an. Die Nahpunktweite ist durch eine leicht verschiebbare, halbtransparente grüne Ebene symbolisiert, deren aktuelle Distanz zur Kamera ebenfalls kontinuierlich angezeigt wird. Das Scheinfenster ist als orange Ebene dargestellt und verfügt, wie die blaue Ebene der Fernpunktweite auch, über eine Distanzanzeige zur Kamera und zu den Ebenen untereinander. Nach der Berechnung der entsprechenden Weiten im neu entwickelten Standardformular für die stereoskopische Produktion, können die entsprechenden Ebenen des Stereo-3D-Tools auf die errechneten Positionen verschoben werden. Anschließend können die Objekte im 3D-Raum, wie beispielsweise die in der Abbildung dargestellte rote Kugel, problemlos in der korrekt geplanten Tiefe im Raum positioniert werden.

Ein weiteres Einsatzgebiet für dieses Tool ist die Erstellung von Animatics zur Planung und Prävisualisierung von Realdrehn. Durch den groben Nachbau des realen Sets in Maya können vor dem eigentlichen Drehbeginn ausgesprochen kostengünstig verschiedene Stereoparameter getestet werden. Dazu gehört beispielsweise die Überprüfung der stereoskopischen Wirkung verschiedener Stereobasen, die Festlegung von Darstellerpositionen oder die Bestimmung zulässiger Aktionsradien von Schauspielern.

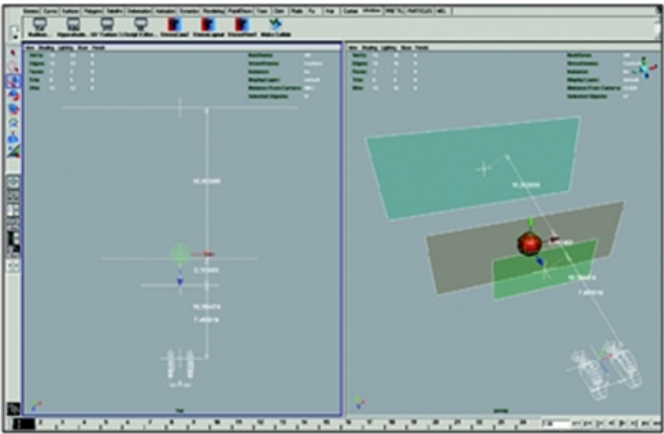


Abbildung. 2.20: Das Stereo-3D-PlanungsTool

2.6 Bewertung des 2D-Fusionsverfahrens

Pro	Contra
<ul style="list-style-type: none">• Bewährte Techniken des digitalen Compositings• Dementsprechend ausgereifte Lösungen am Markt verfügbar.• Schnelle Handhabung der 2D-Elemente.• Echtzeitvorschau je nach Rechnerleistung möglich.• Leichte Montage des stereoskopischen Effekts, da die 2D-Layer leicht gegeneinander verschiebbar sind.• Täuschung der Betrachter ist im 2D-Raum leicht möglich.	<ul style="list-style-type: none">• Keine physikalisch korrekten Schatten.• Keine Interaktion der Schatten mit den 3D-Szenenelementen im Compositing.• Relative Positionierung der Objekte zueinander muss vorgetäuscht werden.

Tabelle 2.2: Bewertung des 2D-Fusionsverfahrens.

3. Alternative Fusionsverfahren

Nachdem im vorherigen Kapitel ein erstes Verfahren zur Fusion von stereoskopisch gedrehten Realaufnahmen eines Darstellers mit einer computergenerierten stereoskopischen Umgebung vorgestellt und untersucht wurde, soll im Folgenden die Frage geklärt werden, welche alternativen Herangehensweisen bei gleichbleibenden Laborbedingungen möglich sind. Neben der Überprüfung der prinzipiellen Machbarkeit, steht dabei das Erfassen der möglichen Vor- und Nachteile der einzelnen Verfahren im Vordergrund des Interesses. Auf Basis dieser gewonnenen Erkenntnisse erfolgt dann im weiteren Verlauf der Arbeit eine Bewertung der identifizierten Fusionsverfahren.

3.1 Darsteller als Textur: 3D-Fusionsverfahren

Während die Zusammenführung des stereoskopischen Contents im Zuge der ersten vorgeschlagenen Lösung komplett im 2D-Raum stattfindet, soll im folgenden alternativen Verfahrensvorschlag die Fusion direkt im 3D-Raum von Maya durchgeführt werden. Dazu muss das real gedrehte Videomaterial des Darstellers als Textur auf einen geometrischen Körper in der computergenerierten Szene projiziert werden. Die Vorarbeiten zu diesem 3D-Fusionsverfahren bleiben bezüglich der notwendigen Maskierungsschritte (Chroma-Keying, Rotoscoping) identisch zum 2D-Fusionsverfahren.

3.1.1 Integration des Darstellers

Um die Integration des Darstellers zu ermöglichen, wurde die bisherige 3D-Umgebung in Maya um eine einfache Polygonfläche - eine sogenannte Plane - erweitert (Create > Polygon Primitive > Plane). Die erstellte Plane, die zur Verdeutlichung des Verfahrens in Abbildung 3.1 dargestellt ist, wurde anhand der festgelegten Dimension einer MayaEinheit auf die Größe des Darstellers skaliert und senkrecht zum Boden aufgestellt. Sie diente im weiteren Verlauf als Träger der Darstellertextur.

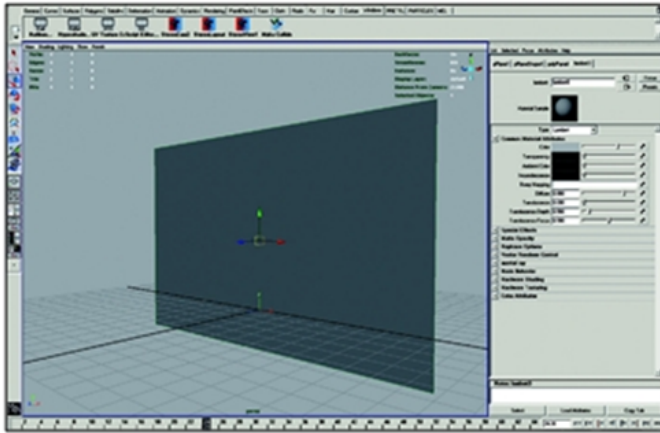


Abbildung. 3.1: Polygonplane als Träger der Darstellertextur

Im nächsten Schritt wurde der Polygonfläche als neues Oberflächenmaterial ein diffuser Lambertshader zugewiesen (Materials > Assign New Material > Lambert). Da dieser Shader nicht spiegelt, können keine unerwünschten Reflexionen oder Glanzpunkte entstehen. Anschließend wurde eine planare Projektionsmethode für die Textur gewählt (Modelling > Polygon UVs > Planar Mapping). Wie der Name bereits andeutet, wird mit dieser Methode eine Textur eben auf die ausgewählte Geometrie projiziert.

Im Attribute Editor des Materials wurde dann der Oberfläche das erste Bild der Darsteller-Bildsequenz aus der linken Ansicht zugewiesen (Color > File). Diese Bildsequenz lag, wie auch beim 2D-Fusionsverfahren, als unkomprimierte Einzelbildfolge im tif.-Format mit integriertem Alphakanal vor. Abbildung 3.2 zeigt die Polygonfläche nach Zuweisung der Darstellertextur. Aufgrund des integrierten Alphakanals ist nur der Darsteller sichtbar - die Polygonfläche selbst wird unsichtbar und dient somit ausschließlich als Träger der Textur. Bei diesem Vorgehen stellte sich heraus, dass die Farbanpassung des Darstellers an den Hintergrund im 2D-Compositing deutlich komfortabler durchzuführen ist.

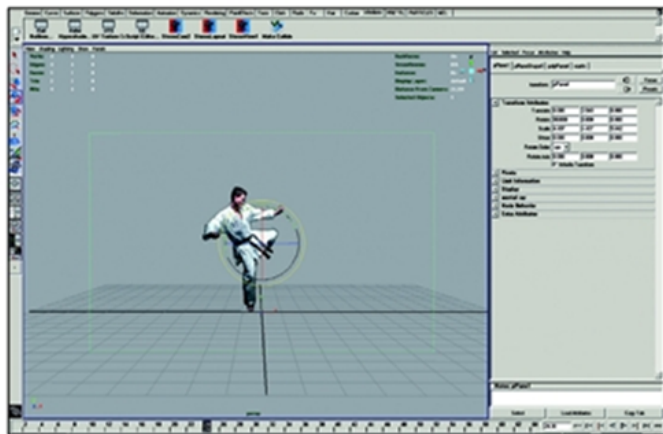


Abbildung. 3.2: Darstellertextur mit Alphakanal auf der Plane

Damit im Laufe der Animation nicht nur das erste Bild gezeigt wird, sondern die gesamte Bildsequenz, muss die Option Use Image Sequence im Attribute Editor des Materials aktiviert werden. Maya lädt damit für jeden Frame der Animation das entsprechende Bild der Sequenz von der Festplatte. Dies funktioniert aber nur dann, wenn die Maya-Namenskonventionen präzise eingehalten werden. Es existieren lediglich drei mögliche Bildbezeichnungformate, mit denen eine Animation aus einer Bildsequenz funktionieren kann (# steht dabei für die Bildnummer): name.#.ext | name.ext.# | name.#.

Es war nun zwingend notwendig, die komplette Animation des Darstellers zu überprüfen, um sicherzustellen, dass alle Bewegungen der kompletten Sequenz auf die Plane passen. Abbildung 3.3 belegt die Problematik anhand eines Frames der Animation, bei dem die Plane nicht mehr ausreichend Platz für die Projektion der Textur bot. Als Folge dessen wurde der Kopf des Darstellers beschnitten.

Diese Schwierigkeit taucht im 2D-Fusionsverfahren nicht auf, da die Bildsequenz des Darstellers als komplette, eigenständige Ebene in voller Auflösung über dem Hintergrund liegt.

Nach Anpassung der Textur wurde die Plane in der gewünschte Tiefe im Bild positioniert. Diese Positionierung ist im 3D-Raum selbstverständlich präziser möglich als im 2D-Raum des ersten Verfahrens, da der Positionierungsvorgang von allen Seiten überprüft werden kann. Damit der Darsteller jedoch optisch richtig auf seinen Füßen stand, musste die Plane nochmals Bild für Bild manuell ausgerichtet werden. Hierbei ist größte Sorgfalt notwendig. Bei einer zu hohen Fussposition, wie sie in Abbildung 3.4 dargestellt ist, wird der Schatten des Darstellers nicht richtig geworfen - der Darsteller

scheint dann zu schweben. Ist der Fuss jedoch zu tief positioniert, dringt er möglicherweise in die Bodengeometrie ein. Beides würde augenblicklich eine realistische Integration zerstören. Gegebenenfalls muss die Polygonfläche über alle drei Achsen per Keyframes von Hand animiert werden, damit sie den Bewegungen des Darstellers folgen kann. Diese relativ zeitbeanspruchende Komplikation tritt im 2D-Raum ebenfalls nicht auf.

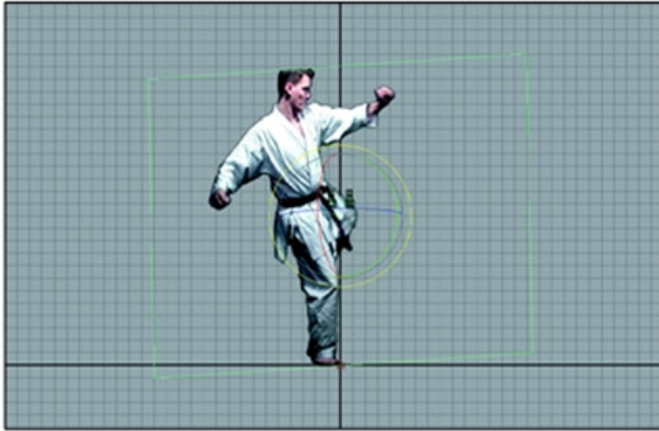


Abbildung. 3.3: Angeschnittener Darsteller während der Animation

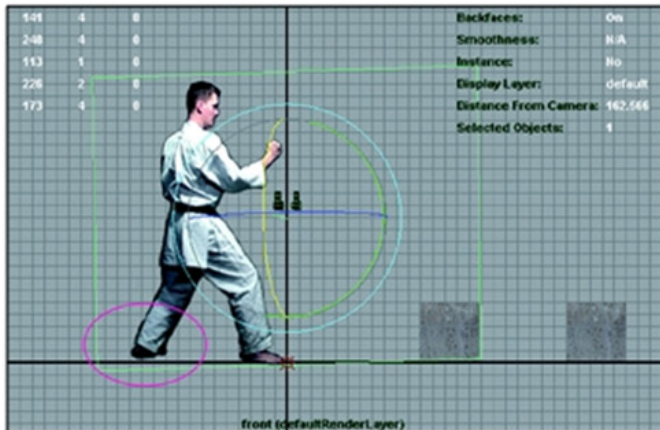


Abbildung. 3.4: Ausrichten der Beine des Darstellers

Nach der Ausrichtung der Plane begannen erste Schattentests. In diesem Bereich liegt offensichtlich einer der größten Vorteile der 3D-Fusion. Denn während bei der 2D-

Fusion der Schatten des Darstellers künstlich als einfache Ebene eingefügt werden muss, wird bei der 3D-Fusion ganz automatisch ein physikalisch korrekter, fein abgestufter Schattenwurf erzeugt, der darüber hinaus auch mit allen Objekten der CG-Szenerie interagiert. Dies funktioniert selbst mit hochkomplexen Objekten wie Nebel oder Rauch. Ein solcher Effekt wäre in einem reinen 2D-Verfahren kaum oder nur mit einem relativ großen Aufwand umzusetzen.

Abbildung 3.5 verdeutlicht die Interaktion anhand eines einfachen Beispielrenderings, bei dem deutlich zu sehen ist, dass der Schatten des Darstellers auf einen dahinterliegenden Steinblockgeworfen wird. Dieser Umstand ist für die räumliche Wirkung einer Szene von großer Bedeutung, da Schatten zu den ausgesprochen wichtigen bildbezogenen monokularen Tiefenreizen gehören. Schatten verdeutlichen die räumliche Beziehung zwischen Objekten innerhalb einer Szene. Sie zeigen, an welcher Stelle ein Objekt den Boden berührt oder aber wie weit ein Objekt sich über dem Boden befindet. Darüber hinaus verstärken Schatten ganz allgemein die dreidimensionale Objektwahrnehmung.



Abbildung. 3.5: Interaktion mit Beispielrendering

3.1.2 Integration der Kugelanimation

Im nächsten Schritt wurde die Animation der explodierenden Steinkugel bereinigt und komplett in die Maya-Szene importiert. (File > Import).

Es erfolgte die Anpassung der Trittbewegung an die Kugel. Dazu mussten Position und Ablauf der Darstellersequenz relativ zur Animation der Steinkugel nivelliert werden.

Während die Position relativ einfach festzulegen war, musste der Animationsablauf komplett überarbeitet und manuell an den Zeitablauf der neuen Szene angepasst werden. Da bei diesem 3D-Fusionsverfahren die Szene inklusive aller Animationen als eine Einheit gerendert wird, ist eine nachträgliche Korrektur des Ablaufs nicht mehr möglich.

3.1.3 Prinzipbedingte Komplikationen

Nachdem die Animation stimmig funktionierte, wurden erste Test-Renderings aus Sicht der linken Kamera angefertigt. Dabei trat eine weitere Problematik deutlich hervor: Während für das 2D-Compositing lediglich eine linke und eine rechte Ansicht des Hintergrundes - Himmel und Gebäudekomplex - gerendert werden mussten, muss bei dem 3D-Ansatz für jedes Einzelbild die komplette Szene einschließlich Darsteller, Steinkugelanimation und Hintergrund komplett neu berechnet werden. Dies dauert in voller HDV-Auflösung von 1440×1080 Pixeln mit Raytracing überproportional lange und erhöht damit die Fehleranfälligkeit beträchtlich. Darüber hinaus ist hierdurch keine Echtzeitvorschau möglich.

In den Testrenderings offenbarten sich weitere prinzipbedingte Komplikationen. Zum einen durchdrang der Darsteller, wie in Abbildung 3.6 dokumentiert ist, mit seinem linken Fuss die Bodengeometrie. Dies war im Ansichtenfenster übersehen worden und erforderte somit eine manuelle Korrektur. Das erneute Rendering dauerte anschließend wieder etwa 15 Minuten pro Frame.



Abbildung. 3.6: Ein Fuss durchdringt die Bodengeometrie



Abbildung. 3.7: Steinsplitter durchdringen die Bodengeometrie

Deutlich problematischer waren jedoch die Auswirkungen der explodierenden Kugel. Die Animation entstand mit Hilfe eines Partikelsystems, weshalb die umherfliegenden Steinbrocken den Gesetzen einer dynamischen Simulation folgen und nicht von Hand zu kontrollieren waren. Dementsprechend kam es auch hier, wie in Abbildung 3.7 zu sehen, zu Durchdringungen der Bodengeometrie. Um dies wirkungsvoll zu verhindern, müssten alle Objekte der Szene in Festkörper - die sogenannten Rigid-Bodies - umgewandelt werden, wodurch sie voneinander abprallen würden, anstatt sich gegenseitig zu durchdringen. Durch die Kollisionseigenschaft der Objekte würde sich aber die Komplexität der Szene drastisch erhöhen, da die Positionen der umherfliegenden Objekte nur noch schwer vorhersagbar wären.

Nach Behebung der Fehler wurde die Szene zunächst aus Sicht der linken Kamera als Einzelbildsequenz gerendert. Anschließend wurde die Darstellertextur der rechten Ansicht auf die Polygonfläche projiziert und die Szene aus Sicht der rechten Kamera gerendert. Die beiden Einzelbildsequenzen wurden dann in After Effects in zwei .avi-Files umgewandelt und im Stereoscopic Player im Anaglyphenverfahren begutachtet.

3.1.4 Bewertung des 3D-Fusionsverfahrens

Pro	Contra
<ul style="list-style-type: none"> • Tatsächlich vorhandene Tiefenausdehnung der Szenenobjekte. • Einfache und präzise Positionierung der Objekte im 3D-Raum möglich. • Physikalisch korrekter Schattenwurf, auch bei hochkomplexen Objekten. • Interaktion der Schatten mit den Szenenobjekten auch bei hochkomplexen Objekten (Nebel, Rauch) problemlos möglich. • Ausgezeichnete Raumwirkung der Kugelanimation (besser als im 2D-Verfahren) 	<ul style="list-style-type: none"> • Jedes Bild der beiden Ansichten muss komplett neu berechnet werden, dadurch exorbitante Renderzeiten. • Keine Echtzeitvorschau möglich. • Hohe Speicherbelastung. • Gravierende Zunahme der Komplexität durch mögliche Kollisionen der Objekte. • Hohe Fehleranfälligkeit bei dynamischen Simulationen. • Aufgrund der physikalisch korrekten Darstellung ist keine Täuschung möglich (z.B. bei Fussposition und Schatten). • Alle Animationskurven müssen einander angepasst werden. • Es werden weniger Layer produziert. Deshalb im Compositing schlechter stereoskopisch zu justieren. • Schlechtere optische Integration des 2D-Darstellers in die 3D-Umgebung. • Es dürfen ausschließlich Raytracing-schatten verwendet werden. • Finale Filmcodierung muss wieder im 2D-Compositing durchgeführt werden.

3.2 Planare Stereoskopie: 2.5D-Fusionsverfahren

Im bisherigen Verlauf der Untersuchung von möglichen Verfahren zur Fusion von stereoskopischen Realaufnahmen und computergenerierten Umgebungen wurde bereits ein 2D- und ein 3D-Fusionsverfahren realisiert und analysiert.

Nachfolgend soll nun ein sogenanntes 2.5D-Verfahren untersucht werden. Dieser Verfahrensansatz unterscheidet sich von einem reinen 2D-Compositing dadurch, dass die zweidimensionalen Bildelemente der einzelnen Ebenen entsprechend ihrer geplanten Lage in negativer z-Richtung in die Tiefe verschoben werden. Durch entsteht bei Kamerabewegungen der Tiefenreiz der Bewegungsparallaxe, woraus ein Mehrwert im Sinne einer besseren Tiefenwirkung resultiert. Inwieweit diesem Vorteil auch Nachteile gegenüberstehen, wird im weiteren Verlauf genauer untersucht.

3.2.1 Ursprünge der 2.5D-Technik

Die 2.5D-Technik etablierte sich innerhalb der Computergrafik immer in den Bereichen, in denen bei möglichst geringer CPU- und Speicherbelastung ein 3D-Eindruck erreicht werden sollte. Als primäres Anwendungsgebiet sind dementsprechend Computerspiele zu nennen, die dem Spieler in Echtzeit eine begehbare 3D-Umgebung anbieten müssen. Um dies zu erreichen, werden in Computerspielen seit Mitte der 70er Jahre des vorigen Jahrhunderts anstatt aufwändiger 3D-Modelle sogenannte Billboards eingesetzt. Ein Billboard (dt.: Plakatwand, Messtafel) ist eine zweidimensionale planare Bildprojektion, die wesentlich schneller zu berechnen ist als ein vergleichbares 3D-Modell.

Ein weiteres populäres Anwendungsgebiet der 2.5D-Technik sind Echtzeitanwendungen wie beispielsweise Landschaftsvisualisierungen, bei denen oft exorbitante Mengen an komplexen Objekten wie Bäume oder Sträucher in Echtzeit dargestellt werden müssen. Eine solche Aufgabe würde bei Verwendung von echten 3D-Modellen die Leistung der meisten Rechner übersteigen. Abbildung 3.8 verdeutlicht die Lösung dieses Problems. Die benötigten Pflanzen werden mittels der Billboardtechnik integriert.

Abbildung 3.9 zeigt die Billboards der Pflanzen von oben, um nochmals den planaren Charakter dieser Technik zu demonstrieren. Die plastische Wirkung dieses Verfahrens kann jedoch relativ einfach verstärkt werden, indem die vorhandene Fläche kopiert und anschließend, um 90° gedreht wird. Mit dieser einfachen Methode, lassen sich selbst ganze Wälder problemlos darstellen.

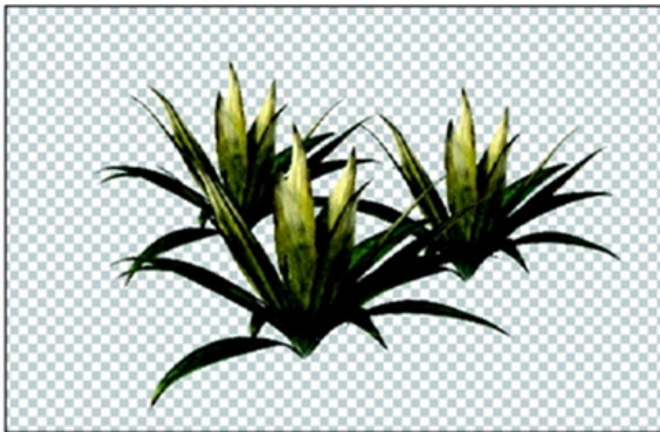


Abbildung. 3.8: Billboarddarstellung von Pflanzen

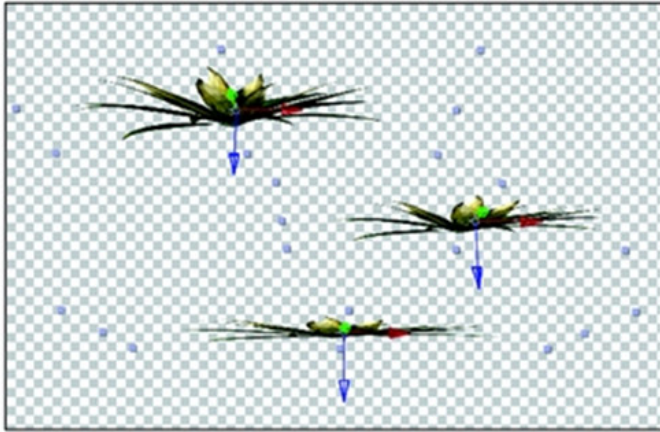


Abbildung. 3.9: Billboarddarstellung der Pflanzen von oben

3.2.2 Prinzipieller Aufbau des 2.5D-Fusionverfahrens

Grundsätzlich betrachtet geht es im Rahmen der vorliegenden Arbeit beim 2.5D-Fusionsverfahren um das Erzeugen eines räumlichen Eindrucks mittels 2D-Bildern, die im simulierten 3D-Raum von After Effects in unterschiedlichen Tiefen senkrecht entlang der z-Achse gestaffelt sind. Dadurch ergeben sich unterschiedliche Distanzen der einzelnen 2D-Bilder zur virtuellen Kamera. Diese Distanzen bewirken bei Kamerabewegungen durch die auftretende Bewegungsparallaxe eine deutlich wahrnehmbare Tiefe. Im Folgenden wird der prinzipielle Aufbau einer Szene im 2.5D-Fusionsverfahren anhand der Beschreibung der Testproduktion verdeutlicht. Um die Stärken des Verfahrens besser demonstrieren zu können, wurden der Szene als neues Vordergrundelement mehrere Pflanzen hinzugefügt.

- Im ersten Schritt wurden alle benötigten Bilddaten in After Effects importiert und eine neue Komposition erstellt.
- Anschließend musste festgelegt werden, welche Objekte in welcher Tiefenebene der Komposition liegen sollten. Im vorliegenden Fall gab es insgesamt sechs verschiedene Tiefenebenen: die beiden Pflanzenebenen im Vordergrund, der Darsteller und die berstende Steinkugel im Mittelgrund, die Häuser in zwei verschiedenen Tiefen im Hintergrund und den Himmel als Ebene mit der größten Tiefe. Die einzelnen Bildelemente wurden entsprechend ihrer späteren Tiefenlokalisation in die Komposition geladen.
- Damit die Ebenen überhaupt in drei Dimensionen zu bearbeiten waren, musste jede einzelne durch Aktivierung des 3D-Würfels im Kompositionsfenster zu einer 3D-Ebene gemacht werden.

- Im folgenden Schritt wurde eine neue Kamera für die linke Ansicht erstellt, mit welcher der spätere Schwenk animiert werden sollte (Ebenen > Neu > Kamera). Als Kameraeinstellung wurde die 35 mm-Vorgabe übernommen.

- Die einzelnen Bildebenen konnten nun entsprechend ihrer zugedachten Tiefe in Richtung der negativen z-Achse verschoben werden. Die Lage der Ebenen kann dabei mit dem Kamera-drehen-Werkzeug kontrolliert werden.

Es ist unbedingt zu beachten, dass die einzelnen Bildelemente nach der Verschiebung in Abhängigkeit ihrer Tiefe skaliert werden müssen. Das bedeutet beispielsweise, dass Vorder- und Mittelgrundobjekte verkleinert werden müssen, während Hintergrundobjekte wie der Himmel deutlich zu vergrößern sind, damit sie weiterhin formatfüllend bleiben. Abbildung 3.11 zeigt den Aufbau der bisherigen Komposition von der Seite, so dass die Verschiebung der einzelnen 2D-Element gut zu sehen ist.

- Der Kameraschwenk wurde anschließend über Positions- und Zielpunkt-Keyframes animiert. Dabei musste darauf geachtet werden, dass während der Bewegung keine Lücken zwischen den einzelnen Elementen von der Kamera erfasst wurden.

- Zuletzt wurde die linke Kamera dupliziert und als rechte Kamera um 20 Pixel nach rechts verschoben. Anschließend konnten die beiden Ansichten gerendert werden.

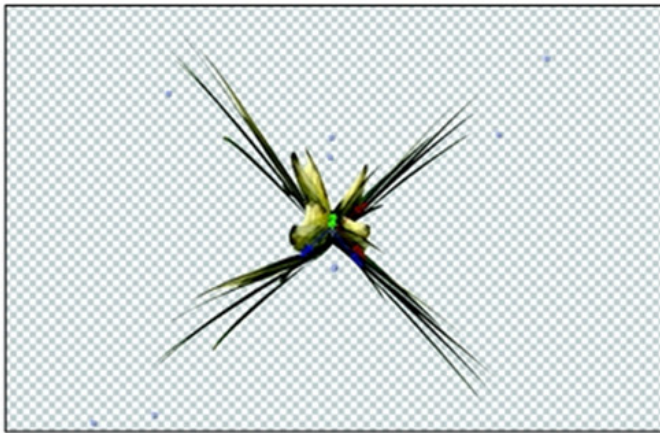


Abbildung. 3.10: Erweitertes Billboarding einer Pflanze

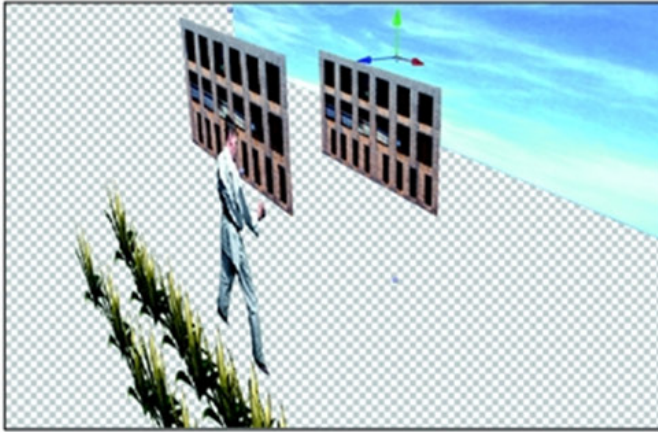


Abbildung. 3.11: Aufbau der 2.5D-Szene entlang der z-Achse

Bei der vorgestellten Verfahrensumsetzung im 2.5D-Raum wurden zwei virtuelle Kameras für die linke und rechte Ansicht generiert. Dies bedeutet, dass der Effekt auch aus konventionellen, monoskopischen Aufnahmen des Darstellers gewonnen werden könnte.

Der Vollständigkeit halber sei an dieser Stelle erwähnt, dass es mit dem 2.5D-Verfahren sogar möglich ist, einen Stereo-3D-Film aus einem einfachen 2D-Film zu erstellen. Dazu müssen alle freistehenden Elemente der Szene Frame für Frame aufwändig per Rotoskoping manuell extrahiert und im 3D-Raum wieder auf unterschiedliche Distanzen zur virtuellen Kamera gebracht werden. Nachdem ein Objekt freigestellt und in eine neue Tiefenebene verschoben ist, bleibt im Originalbild, welches jetzt als Hintergrund dient, ein Loch zurück. Es fehlen somit Bildinformationen, die ersetzt werden müssen. Um dies zu erreichen gibt es mehrere Methoden, angefangen von hoch entwickelten Plug-ins für Pattern Rekonstruktion¹ bis hin zu einfachem Klonen oder Verwischen der angrenzenden Hintergrundteilbereiche.

Die stereoskopischen Ergebnisse dieses Verfahrens können durchaus beeindruckend sein, denn schließlich entsteht durch die starke Bewegungsparallaxe bei Kamerabewegungen eine sehr gute räumliche Tiefenwirkung. Die praktische Durchführung des 2.5D-Fusionsverfahrens hat gezeigt, dass dies insbesondere dann gilt, wenn viele Ebenen unterschiedlicher Tiefe in der Szene zur Verfügung stehen, da in diesem Fall viele parallaktische Verschiebungen gegeneinander stattfinden. Ohne die beiden Pflanzenebenen im Vordergrund ist der stereoskopische Effekt weit weniger stark ausgeprägt als mit den Pflanzen. Zur Verdeutlichung der verstärkten Tiefenwirkung durch die Pflanzen zeigt Abbildung 3.12 ein Standbild aus dem 2.5D-Fusionsverfahren.

Ein weiterer wesentlicher Vorteil, der sich aus einer großen Anzahl vorhandener Bildebenen ergibt, ist die Möglichkeit, die einzelnen Layer problemlos auf verschiedene Tiefen einzustellen, was eine präzise Kontrolle über die gleichmäßige Tiefenwirkung der gesamten Szene zulässt.

Der prinzipbedingte größte Vorteil des 2.5D-Fusionsverfahrens gegenüber dem 3D-Verfahren ist die sehr kurze Renderzeit bei geringer Speicherbelastung, wodurch sogar eine Echtzeitvorschau möglich wird.

Abbildung 3.13 veranschaulicht einen bedeutenden Vorteil, den das 2.5D-Fusionsverfahren gegenüber dem 2D-Verfahren bietet. Durch die in die Tiefe gestaffelten Bildebenen tritt der Tiefenreiz der Verdeckung verstärkt in den Vordergrund. Darüber hinaus ist im 2.5D-Raum ein realistischerer Schattenwurf mittels virtueller Lichtquellen möglich.



Abbildung. 3.12: Stereoskopisches Ergebnis des 2.5D-Fusionsverfahrens

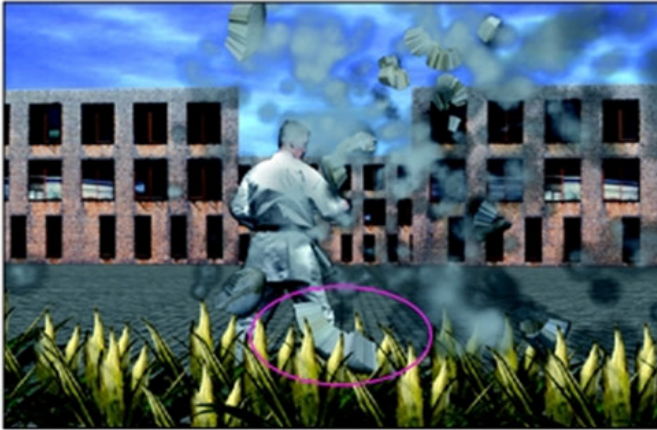


Abbildung. 3.13: Verdeckung als positiver Effekt des 2.5D-Fusionsverfahrens

3.2.3 Prinzipbedingte Komplikationen

Die konkrete Umsetzung der Testproduktion im 2.5D-Fusionsverfahren unterschied sich in einigen Punkten deutlich von den bisher untersuchten Methoden, da die verwendeten 2D-Bilder gewisse Bedingungen erfüllen müssen, damit der Tiefeneffekt zufriedenstellend zustande kommt. Zu diesen Bedingungen gehört beispielsweise, dass in den verwendeten 2D-Bildern selbst keine Perspektive und kein Schatten vorhanden sein darf, da diese sich bei Kamerabewegungen selbstverständlich nicht entsprechend anpassen - und somit die Tiefenwahrnehmung zerstören würden. Aus diesem Grund konnten die bisher genutzten gerenderten Hintergrundbilder aus Maya nicht mehr verwendet werden, weshalb auf die Originaltexturen zurückgegriffen werden musste.

Der größte prinzipbedingte Nachteil des 2.5D-Fusionsverfahrens ist die fehlende Tiefenwirkung der Objekte selbst. In diesem Zusammenhang wird oft vom sogenannten cardboard-look gesprochen. Mit diesem Begriff wird das Aussehen der 2D-Bildelemente mit Pappaufstellern verglichen. Abbildung 3.14 veranschaulicht den cardboard-look.



Abbildung. 3.14: Cardboard-look des 2.5D-Fusionsverfahrens



Abbildung. 3.15: Sichtbare Ebenenbegrenzung im 2.5D-Raum

Die Animation der berstenden Steinkugel erschwerte die Zusammensetzung der Sequenz erheblich. Grund dafür war die formatfüllende Ausdehnung der Steinsplinter über die gesamte Ebene. Wie in Abbildung 3.15 dokumentiert ist, waren dadurch die Begrenzungen der Ebene ab einem gewissen Kamerawinkel zu sehen. Der Spielraum für Kamerabewegungen ist also im Vergleich zum 3D-Raum erheblich eingeschränkt. Ein weiterer problematischer Punkt war das Bodenelement. Während dessen Verwendung weder im 2D-Raum noch im 3D-Raum zu Problemen führte, verschob es sich im 2.5D-Raum aufgrund der größeren Entfernung zur virtuellen Kamera in Relation zum

Darsteller langsamer. Dadurch schien der Darsteller über den Boden zu schweben.

Abhilfe schafft hier eine rein orthogonale Ansicht von der Seite. Dadurch wird die störende Perspektive des Bodenelements eliminiert. Eine bessere Methode, diese Problematik zu umgehen, ist - wie im vorliegenden Fall - die Nutzung von Vordergrundelementen, die den Problembereich kaschieren und darüber hinaus auch noch den Tiefeneindruck der Szene verstärken.

3.2.4 Bewertung des 2.5D-Fusionsverfahrens

Pro	Contra
<ul style="list-style-type: none"> • Starke Bewegungsparallaxe bei Kamerabewegungen möglich, dadurch sehr gute räumliche Tiefenwirkung. • Extrem große Kontrolle über die Tiefenwirkung, da viele Layer zur Verfügung stehen, die sehr exakt manipuliert werden können. • Im Vergleich zum 3D-Verfahren sehr kurze Renderzeit. • Im Gegensatz zum 3D-Verfahren sehr geringe Speicherbelastung. • Echtzeitvorschau möglich. • Im Gegensatz zum 2D-Verfahren ist ein realistischer Schattenwurf mittels virtueller Lichtquellen möglich. 	<ul style="list-style-type: none"> • Keine echte Tiefe der Elemente. • Flache 3D-Wirkung der Objekte, typischer „cardboard-look“. • Extrem hoher Maskieraufwand. • Ausgangstexturen dürfen keine eigene Perspektive besitzen • Schattenwurf kann die fehlende Tiefe offenbaren. • Im Gegensatz zum 3D-Verfahren geringer Spielraum für Kamerabewegungen. Die Grenzen der 2D-Ebenen könnten sichtbar werden. • Keine echte Integration von hochkomplexen Objekten möglich.

Tabelle 3.2: Bewertung des 2.5D-Fusionsverfahrens.

4. Diskussion der Ergebnisse

Alle Vorzüge und Problematiken der jeweiligen identifizierten Fusionsverfahren sollen abschliessend nochmals kritisch gegeneinander abgewogen werden. Es soll die Frage beantwortet werden, ob es das ideale Verfahren, das allen Produktionsanforderungen gewachsen ist, überhaupt gibt - oder ob es nicht sinnvoller ist, jeder Aufnahmesituation die effektivste Kombination der identifizierten stereoskopische Fusionsverfahren zuzuordnen.

4.1 Ausgangssituation

Im Verlauf der Untersuchungen wurden drei verschiedene Verfahren zur Fusion von stereoskopisch gedrehten Realaufnahmen mit stereoskopisch gerenderten CG-Umgebungen identifiziert. Deren praktische Durchführbarkeit und Effizienz wurden mittels einer Testproduktion nachgewiesen und ausführlich erprobt.

Im Folgenden soll nun die entscheidende finale Frage diskutiert werden, welches der drei Fusionserfahren den Produktionsanforderungen am ehesten gerecht wird. Um dies fundiert bewerkstelligen zu können, werden die entsprechenden Anforderungen im Folgenden zunächst nochmals explizit rekapituliert.

4.1.1 Anforderungen an die visuellen Integrationsoptionen

Wie im ersten Kapitel ausführlich beschrieben, kommt es bei einem Compositing primär darauf an, den Zuschauer möglichst nicht erkennen zu lassen, dass das betrachtete Filmbild aus mehreren verschiedenen Einzelementen besteht. Für den gewünschten homogenen Gesamteindruck der Szene ist vor allem die Qualität des Keyings entscheidend. Ein mangelhaft ausgeführtes Keying führt im finalen Compositing zu sogenannten Matte Lines, die vom Betrachter als helle oder dunkle Linie um ein maskiertes Objekt wahrgenommen werden können, wodurch die visuelle Integration des entsprechenden Objekts in das Gesamtbild nicht mehr sauber gelingen kann. Ein weiteres Phänomen, welches die Integration maskierter Objekte deutlich erschwert, ist das sogenannte Color Spill, eine Farbüberstrahlung des Blue- oder Greenscreens auf hellen Objekte, die gekeyt werden sollen.

In der konkreten Testproduktion erschwerten die ungünstigen Bedingungen beim Bluescreen-Dreh und die schlechte Aufnahmequalität das Keying erheblich, wodurch der Schauspieler nicht sauber maskiert werden konnte und Matte Lines entstanden. Darüber hinaus kam es aufgrund des weissen Karate-Anzugs zu einem erheblichen Color Spill.

Aufgrund dieser möglichen Problematiken ist es essentiell, dass ein Fusionsverfahren ausreichend gestalterische Optionen bietet, die beim Compositingprozess zur Verbesserung der visuellen Integration genutzt werden können. Zu diesen Optionen gehören in erster Linie Techniken zur Maskenoptimierung und zur Farbanpassung.

4.1.2 Anforderungen an das Interaktionspotential

Unter dem Interaktionspotential wird im vorliegenden Zusammenhang der Umfang der Beeinflussungsmöglichkeiten verstanden, die das jeweilige Fusionsverfahren zwischen den verschiedenen Objekten der Szene zulässt.

Zu den wichtigsten Eigenschaften, die eine Interaktion zwischen den Szenenobjekten verdeutlichen, zählen Schattenwurf und Schattenempfang, sowie die Kollisionsfähigkeit. Dies lässt sich damit begründen, dass Schatten und Kollisionen Informationen über die relativen Positionen der Vorder- und Hintergrundobjekte zueinander liefern, was die Lokalisation der einzelnen Objektiefen ermöglicht.

Eine derartige Interaktion trägt beispielsweise entscheidend zur glaubwürdigen Integration des realen Darstellers in die CG-Umgebung bei. Dementsprechend ist die Möglichkeit zur Schaffung und Bearbeitung solcher Interaktionseffekte eine weitere wesentliche Anforderung an die jeweiligen Fusionsverfahren.

4.1.3 Anforderungen an die stereoskopische Justierbarkeit

Ein weiteres wichtiges Kriterium für die Anwendbarkeit der einzelnen Fusionsverfahren ist die Justierbarkeit der stereoskopischen Tiefenwirkung beim Compositingvorgang.

Aufgrund der Korrelation zwischen der Position der korrespondierenden Bildpunkte und der wahrnehmbaren Stereotiefe ist es möglich, durch Verschiebung der Stereo-Halbbilder zu bestimmen, in welcher Tiefe die einzelnen Elemente der Szene erscheinen werden.

Diese Tatsache ist für den Compositingprozess entscheidend, bei dem schließlich mehrere unterschiedliche Elemente zu einem stimmigen Stereo-3D-Bild mit einer durchgehenden Tiefe zusammengesetzt werden müssen. Das bedeutet konkret, dass die beiden Ansichten jedes einzelnen Objekts so auszurichten sind, dass die daraus resultierende wahrnehmbare Tiefe mit der theoretischen Position des Objekts in der Szene übereinstimmt.

Eine solch präzise Justierung der stereoskopischen Tiefe wird aber nur dann möglich sein, wenn das gewählten Fusionsverfahren es prinzipiell zulässt, die einzelnen Szeneobjekte auf jeweils separaten Ebenen zu verwenden.

4.1.4 Anforderungen an die Effizienz

Bei der Analyse möglicher Fusionsverfahren war nicht allein die grundsätzliche Machbarkeit entscheidend, sondern auch die Effizienz, denn schließlich sollten die vorgeschlagenen Verfahren an regulären Desktop-Rechnern mit der an der Fakultät vorhandenen Software durchführbar sein. Dies führte zwangsläufig zu einer effizienzbedingten Restriktion bei der Durchführung der Fusionsverfahren.

Zu den Anforderungen an die Effizienz zählt zunächst die Einhaltung einer möglichst

geringen Renderzeit, was sich darauf zurückführen lässt, dass eventuell nicht für alle produktionen eine Renderfarm zur umfangreichen Berechnung komplexer CG-Szenarien zur Verfügung steht.

4.2 Diskussion

Im Folgenden werden die wichtigsten identifizierten Vorzüge und Problematiken der jeweiligen vorgeschlagenen Fusionsverfahren nochmals kritisch gegeneinander abgewogen, wobei besonderer Wert auf die Umsetzbarkeit der vorab definierten Anforderungen gelegt wird.

Dabei werden die Compositingelemente Schauspieler, Steinkugel und Hintergrundgebäude aufgrund ihrer jeweiligen speziellen Eigenheiten im weiteren Verlauf der Untersuchung bei Bedarf getrennt betrachtet und bewertet.

4.2.1 Die visuellen Integrationsoptionen der Fusionsverfahren

Der größte Vorteil des 2D-Fusionsverfahrens beruht auf der Tatsache, dass es ausschließlich auf bereits bewährten Techniken des zweidimensionalen digitalen Compositings basiert. Diese Techniken werden seit Jahren weiterentwickelt und sind dementsprechend ausgereift. Speziell im Bereich der Maskenoptimierung und der Farbkorrektur stehen ausgesprochen mächtige Werkzeuge zur Verfügung, was sich vor allem bei der Integration des gekeyten Darstellers in den CG-Hintergrund bemerkbar macht. Mit Hilfe hochspezialisierter Werkzeugoptionen können einerseits Matte Lines und Color Spill in Abhängigkeit von der Qualität des Ausgangsmaterials reduziert oder vollständig entfernt werden und andererseits Vordergrundobjekte wie der Darsteller oder die Animation der berstenden Steinkugel an den Hintergrund farblich präzise angepasst werden.

Das 3D-Fusionsverfahren kann die Anforderungen in diesem Bereich bei weitem nicht in diesem Umfang erfüllen. Die Maskierung erfolgt zwar prinzipiell auf die gleiche Weise wie beim 2D-Fusionsverfahren, die Integration des 2D-Darstellers als Textur in den 3D-Hintergrund ist aufgrund fehlender Anpassungswerkzeuge aber ungleich problematischer, wodurch die visuelle Integration des 2D-Darstellers in die CG-Umgebung qualitativ deutlich schlechter ausfällt als beim 2D-Fusionsverfahren.

Bei der berstenden Steinkugel kommt dagegen der größte Vorteil des 3D-Fusionsverfahrens zum Tragen, da diese Animation eben nicht wie bei den beiden anderen Verfahren Bild für Bild integriert werden muss, sondern vielmehr ein integraler Bestandteil der Szene selbst ist. Dadurch stimmt selbstverständlich sowohl die Beleuchtung als auch die Farbstimmung perfekt mit den übrigen Objekten überein.

Das 2.5D-Fusionsverfahren basiert auf denselben Basistechniken wie das 2D-Verfahren¹, könnte also die visuelle Integration des Darstellers und der Steinkugel auf gleichem qualitativen Niveau erreichen. Durch die Verschiebung der Szenenobjekte entlang der z-Achse werden jedoch die Umrisskanten der einzelnen Bildelemente ver-

stärkt wahrgenommen. Dieses Problem tritt insbesondere bei Kameraschwenks auf, da in diesem Fall ein typischer cardboard-look entsteht, der die Flachheit der Objekte verdeutlicht und damit die visuelle Integration erschwert.

4.2.2 Das Interaktionspotential der Fusionsverfahren

Bei den Anforderungen an die Interaktionsmöglichkeiten offenbart das 2D-Fusionsverfahren seine Schwächen. Im Gegensatz zum 3D-Verfahren werden hier von den Szenenobjekten keine physikalisch korrekten Schatten geworfen. Soll der Darsteller also einen Schatten auf den CG-Hintergrund werfen, muss dieser entweder im Compositing aufwändig simuliert oder aber alternativ beim Dreh vor Bluescreen mit aufgezeichnet und maskiert werden. Besonders problematisch wird es dann, wenn der Darsteller nicht nur einen Schatten werfen soll, sondern die CG-Elemente diesen auch empfangen müssen. In diesem Fall stößt das 2D-Fusionsverfahren aufgrund des Arbeitsaufwands an seine Grenzen.

Ein weiteres gravierendes Problem stellt in diesem Zusammenhang die gewünschte Kollision von Szenenobjekten untereinander dar. Da beim 2D-Fusionsverfahren die zweidimensionalen Bildelemente in Ebenen direkt übereinander liegen, können Elemente auf Vordergrundebenen niemals mit Elementen auf Hintergrundebenen kollidieren oder von diesen verdeckt werden - was gegebenenfalls völlig unrealistisch wirkt. Bei der konkreten Testproduktion wäre es somit nur mit erheblichem Aufwand möglich, dass beispielsweise die Trümmer der geborstenen Steinkugel, die auf der obersten Bildebene positioniert sind, von dem CG-Boden, der sich auf der untersten Ebene befindet, abprallen.

Das 3D-Fusionsverfahren ist im Gegensatz dazu zweifellos für die Objektinteraktion prädestiniert. Der Schattenwurf muss nicht umständlich simuliert werden, sondern wird automatisch und physikalisch korrekt erzeugt. Die Interaktion der Schatten mit allen Szenenobjekten ist selbst bei hochkomplexen Objekten wie den Trümmern und dem Explosionsstaub problemlos möglich.

Darüber hinaus besteht beim 3D-Fusionsverfahren die Möglichkeit, alle Objekte der Szene in sogenannte Festkörper umzuwandeln. Dies hätte zur Folge, dass umherfliegende Steinbrocken andere Objekte wie beispielsweise den Boden, nicht mehr einfach durchdringen, sondern automatisch von ihnen abprallen würden. Ein solches realistisches Verhalten ist mit keinem der anderen Verfahren umsetzbar.

Die Interaktionsmöglichkeiten des 2.5D-Fusionsverfahrens bewegen sich zwischen denen des 2D- und des 3D-Verfahrens. Aufgrund der Verschiebung der Szenenobjekte entlang der z-Achse, ergibt sich eine tatsächliche Tiefe, durch die ein realistischer Schattenwurf mittels virtueller Lichtquellen möglich wird. Außerdem können durch die Tiefendistanzen einzelne Objekte problemlos zwischen den übrigen Szenenobjekten hindurch animiert werden. Eine realistische Kollisionsdetektion wie beim 3D-Fusionsverfahren ist jedoch nicht möglich.

4.2.3 Die stereoskopische Justierbarkeit der Fusionsverfahren

Sowohl das 2D- als auch das 2.5D-Fusionsverfahren ermöglichen dem Benutzer eine leichte und komfortable Justierung des stereoskopischen Effekts, da sich die zahlreich vorhandenen 2D-Ebenen leicht gegeneinander verschieben lassen.

Im Gegensatz dazu besteht ein gerendertes Stereo-Halbbild im 3D-Fusionsverfahren aus einer einzigen Ebene. Dadurch können die beiden Ansichten nur noch im Ganzen justiert werden - eine präzise Tiefeneinstellung einzelner Bildelemente ist somit nicht mehr möglich.

4.2.4 Die Effizienz der Fusionsverfahren

Das vorgeschlagene 2D-Fusionsverfahren ermöglicht prinzipiell eine schnelle Handhabung und Justage der verwendeten zweidimensionalen Bildelemente. Eine komfortable Echtzeitvorschau der Szenenanimation ist aufgrund der Rechenleistung aktueller Desktop-Computer in der Regel problemlos möglich.

Dies gilt grundsätzlich auch für das 2.5D-Verfahren, solange nicht überproportional viele virtuelle Lichtquellen und Schatten verwendet werden.

Demgegenüber offenbart das 3D-Verfahren im Bereich der Effizienz seine größte Schwäche. Die komplette Neuberechnung jedes Einzelbildes der beiden Ansichten hat eine exorbitante Renderzeit zur Folge. Beim 2D-Verfahren müssen beispielsweise die Hintergrundgebäude nur einmal für jede Ansicht in Maya gerendert werden und können anschließend in After Effects als einfache Hintergrundbilder verwendet werden. Bei der 3D-Fusion müssen die Gebäude dagegen für jeden Frame unter Berücksichtigung aller Lichtquellen vollkommen neu berechnet werden. Die daraus resultierende CPU- und Speicherbelastung fordert das verwendete Rechnersystem massiv. Eine Echtzeitvorschau ist damit selbstverständlich ausgeschlossen.

4.3 Ergebnisse der Diskussion

Es kann offensichtlich festgestellt werden, dass keines der drei identifizierten Fusionsverfahren allen gestellten Anforderungen vollkommen gerecht wird.

Während das 2D-Verfahren auf die besten Werkzeuge zur visuellen Integration zurückgreifen kann und sowohl bei der stereoskopischen Justierbarkeit als auch bei der Effizienz punktet, offenbart es bei den Interaktionsmöglichkeiten gravierende Schwächen.

Das Objektinteraktionspotential ist dagegen die größte Stärke des 3D-Fusionsverfahrens. Kein anderes Verfahren kann physikalisch korrekte Schatten dieser Qualität automatisch erzeugen, geschweige denn eine Kollisionsdetektion anbieten. Dies geschieht jedoch auf Kosten der Effizienz und der finalen Tiefenjustierbarkeit. Darüber hinaus stehen nur rudimentäre visuelle Integrationsoptionen zur Verfügung.

Der Vorteil des 2.5D-Fusionsverfahrens gegenüber dem 2D-Verfahren liegt in der tat-

sächlich vorhandenen Tiefenausdehnung der Szene. Bei Kameranähen hat dies eine starke Bewegungsparrallaxe zur Folge, welche die stereoskopische Tiefenwirkung deutlich verstärkt - allerdings auch die Flachheit der einzelnen Bildelemente erhöhen lässt.

Integrationsmöglichkeit, Effizienz und stereoskopische Justierbarkeit sind aufgrund des prinzipiell ähnlichen zweidimensionalen Szenenaufbaus weitestgehend mit denen des 2D-Verfahrens identisch.

Die Interaktionsmöglichkeiten sind im Vergleich zum 3D-Fusionsverfahren jedoch stark eingeschränkt. Einfache Schatten sind zwar grundsätzlich möglich, können aber qualitativ keinesfalls mit denen des 3D-Verfahrens konkurrieren. Darüber hinaus ist keine Kollisionsdetektion verfügbar.

4.4 Fazit

Aus der vorangegangenen Diskussion lässt sich die evidente Erkenntnis ableiten, dass es das ideale stereoskopische Fusionsverfahren, das allen gestellten Produktionsanforderungen uneingeschränkt gewachsen ist, offensichtlich nicht gibt.

Vielmehr ist davon auszugehen, dass jeder geplanten Produktionssituation im Vorfeld die effektivste Kombination der identifizierten Fusionsverfahren zugeordnet werden muss. Entscheidend für eine qualitativ hochwertige, homogene Fusion ist letztendlich also die richtige Mischung aus den drei untersuchten Verfahren.

Für die konkrete Verfahrensdurchführung bedeutet dies, dass eine entsprechende Szene in der Planungsphase zunächst kognitiv in ihre einzelnen Bestandteile zu zerlegen ist. Anschließend muss jedem dieser Bildbestandteile ein adäquates Fusionsverfahren zugewiesen werden, welches dann im Verlauf der Postproduktion umgesetzt wird. Vor dem Hintergrund dieser Erkenntnis, würde sich für die in der vorliegenden Arbeit verwendete Testproduktion folgende Verfahrenszuordnung anbieten:

- Das Himmelelement im Hintergrund reicht theoretisch in die optische Unendlichkeit und hat deshalb kaum eine binokulare Bedeutung. Es kann somit problemlos mittels des 2D-Verfahrens in die Szene integriert werden.
- Die Hintergrundgebäude müssen aufgrund der Tatsache, dass sie keine Schatten von einem Vordergrundobjekt empfangen, nur einmal pro Ansicht in höchster Qualität gerendert werden und können anschließend ebenfalls ganz legitim mittels des 2D-Verfahrens in die Szene übernommen werden.
- Die Animation der berstenden Steinkugel stellt mit Abstand die höchsten Anforderungen an ein mögliches Fusionsverfahren. Aufgrund der immens hohen

Anzahl an schattenwerfenden Trümmerstücken bietet das 3D-Fusionsverfahren die einzig ernsthafte Alternative. Darüber hinaus kann bei diesem Verfahren das Bodenobjekt zu einem Festkörper gewandelt werden, auf dem die Bruchstücke realistisch abprallen können. Die Umsetzung erfolgt durch das Rendern der Animation und des Bodens aus beiden Ansichten. Die entstandenen Einzelbildsequenzen werden dann im 2D-Verfahren der Gesamtszene hinzugefügt.

- Für den Darsteller bietet sich bei der vorliegenden Testproduktion eine zweidimensionale Integration an. Dies kann damit begründet werden, dass dessen Schatten aufgrund des ebenen Hintergrunds ebenfalls problemlos als 2D-Element auf einer neuen Ebene angelegt werden kann. Sollte der Schattenwurf auf geometrisch einfache Hintergrundobjekte fallen, könnten diese bei den Greenscreen-Aufnahmen in grün nachgebaut werden und der resultierende Schatten mit maskiert werden. Sollten die Hintergrundobjekte, auf die der Schatten des Darstellers fallen würde, zu komplex sein, bleibt als Lösung jedoch nur eine zeit- und rechenaufwändige 3D-Fusion.

- Die Integration der Pflanzen, die sich extrem nah im Vordergrund befinden, sollte idealerweise mittels des 2.5D-Fusionsverfahrens durchgeführt werden, da dadurch bei Kamerabewegungen der stereoskopische Tiefenendruck nochmals verstärkt werden kann.

Die Verwendung einer Verfahrenskombination zur Fusion von stereoskopisch gedrehten Realaufnahmen mit stereoskopisch gerenderten CG-Umgebungen sichert ein ausgewogenes Verhältnis von Qualität und Effizienz, bringt aber durchaus auch einen intensiveren Planungsaufwand und neue potentielle Fehlerquellen mit sich.

Die auf eine determinierte Szene optimal abgestimmte Auswahl und Zuordnung der entsprechenden Fusionsverfahren ist dementsprechend ein aufwändiger und komplexer Prozess, der ein fundiertes Hintergrundwissen in den Bereichen Stereoskopie und Postproduktion erfordert.

Literaturverzeichnis

- Alias Wavefront 2004: Learning Maya 6. MEL Fundamentals; a hands-on introduction to Maya Embedded Language. Toronto: Maya.
- Allen, Damian 2007: Encyclopedia of visual effects. Berkeley, Calif., London: Peachpit Press; Pearson Education.
- Bär, Gert 2001: Geometrie - Eine Einführung für Ingenieure und Naturwissenschaftler. Teubner.
- Bahr, Achim 1993: Anaglyphen. In: 3D-Magazin 4, Haltern. Engl. Fassung: Anaglyphs. In: Stereoscopy Nr. 2/36, International Stereoscopic Union 1998.
- Barlow, H. B. & Blamore C. & Pettigrew, J. D. 1967: The neural mechanism of binocular depth discrimination. In: J. Physiol 193, S.327-342
- Bartels, Heinz 1995: Physiologie. Lehrbuch und Atlas; 20 Tabellen. 5., überarb. Aufl. München, Wien, Baltimore: Urban und Schwarzenberg.
- Bender, Michael & Brill, Manfred 2006: Computergrafik. Ein anwendungsorientiertes Lehrbuch. 2., überarb. Aufl. München: Hanser.
- Benninghoff, Alfred 1994: Anatomie, Bd. 1: Zellenlehre und Gewebelehre, Entwicklungsbiologie, Bewegungsapparat, Herz-Kreislauf-System, Immunsystem, Atemapparat: 15., völlig neu bearb. Aufl. München: Urban & Schwarzenberg.
- Benninghoff, Alfred 1994: Anatomie, Bd. 2: Niere, Reproduktionsorgane, endokrine Drüsen, Nervensystem, Sinnesorgane, Haut. 15., völlig neu bearb. Aufl. München: Urban & Schwarzenberg.
- Bertram, Sacha 2005: VFX. Konstanz: UVK-Verl.-Ges. (Praxis Film, 22).
- Birbaumer, N. & Schmidt, R. F. 2006: Biologische Psychologie. 6. Auflage. Berlin, Heidelberg: Springer Medizin Verlag Heidelberg.
- Birn, Jeremy 2006: Digital lighting & rendering. 2. ed. Berkeley, Calif.: New Riders.
- Blake, R. & Hirsch, H. V. B. 1975: Deficits in binocular depth perception in cats after alternating monocular deprivation. In: Science 190, S. 1114-1116
- Bortz, Jürgen & Döring, Nicola 2006: Forschungsmethoden und Evaluation. Für Human- und Sozialwissenschaftler. 4., überarbeitete Auflage. Berlin, Heidelberg: Springer Medizin Verlag Heidelberg.
- Bowers, Brian 1975: Sir Charles Wheatstone FRS, 1802-1875. Science Museum (Hg.). London: H.M.S.O.
- Bowers, Brian 2001: Sir Charles Wheatstone FRS. 1802-1875. 2. ed. London: IEE (IEE history of technology series, 29).
- Choi, Jae-jin 2004: Maya. Character animation, 2nd edition. Sophisticated modeling and animation techniques. 2nd ed. Alameda CA: SYBEX.
- Churchland, P. S. & Ramachandran, V. S. 1996: Filling in: Why Dennett is wrong. In
- K. Atkins (Hrsg.), Perception. Oxford, England: Oxford Press

- Cohen, Michael F. & Wallace, John R. 1993: Radiosiy and Realistic Image Synthesis. Academic Press, 1993
- Crone, Robert A. 1992: The history of stereoscopy. In: Documenta Ophthalmologica 81.
- The official journal of the International Society for Clinical Electrophysiology of Vision, S. 1-16
- Cunningham, Andrew 1997: The anatomical renaissance. The resurrection of the anatomical projects of the ancients. Aldershot: Scolar Press [u.a.].
- Cutting, J.E. & Vishton, P.M. 1995: Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In Epstein, W. & Rogers, S. (Hrsg.), Handbook of perception and cognition: Perception of space and motion, S.69-117. New York: Academic Press.
- Descartes, René 1637: Discours de la méthode pour bien conduire sa raison et chercher la vérité dans les sciences (dt: „Abhandlung über die Methode des richtigen Vernunftgebrauchs und der wissenschaftlichen Wahrheitsforschung“). Eine Übersetzung des Textes ist online verfügbar unter: <http://www.textlog.de/descartes-methode.html>, zuletzt geprüft am 01.04.2008.
- Driemeyer, Thomas & Herken, Rolf 2005: Rendering with mental ray, mental ray Handbooks Vol. 1. 3. völlig überarbeitete Ausgabe. Springer-Verlag. Wien, New York.
- Driemeyer, Thomas 2005: Programming mental ray. mental ray Handbooks, Vol. 2. 3. völlig überarbeitete Ausgabe. Springer-Verlag. Wien, New York.
- Epstein, William 1965: Nonrelational judgements of size and distance. In: American Journal of Psychology 78, S. 120-123
- Fahle, M. 1982: The apparent size of objects situated in front of or behind the horopter. In: Graefe's archive for clinical and experimental ophthalmology. Official organ of the Club Jules Gonin - Albrecht von Graefe's Archiv für klinische und experimentelle Ophthalmologie. Berlin, Heidelberg: Springer. Volume 225, Number 1 / Februar 1987.
- Faller, A. & Schünke, M. & Schünke, G. 1995: Der Körper des Menschen. Einführung in Bau und Funktion. 12. Aufl. Stuttgart: Thieme [u.a.]
- Fellner, Wolf-Dietrich 1988: Computer-Grafik. Mannheim: BI-Wissenschaftsverl. (Reihe Informatik, 58).
- Fleischer, F. P. & Ludwig, W. & Wartmann, R. 1952: Der optimale Horopter. In: Pflügers Archiv, Bd. 254, S. 377-392
- Fletcher, Ni 2007: Highway. Das JVC Kundenmagazin. Chefredakteur: Nick-Fletcher. JVC Professional Europe Ltd. Ausgabe September 2007
- Foley, Jim & Dam, Andries van & Feiner, Steven & Hughes, John 1991: Computer Graphics - Principles and Practice. Addison-Wesley.
- Frank, Anja 2007: Bachelor-Thesis: Stereoskopische High Definition Produktion. Hochschule Furtwangen University, Fakultät Digitale Medien, Sommersemester 2007

- Fries, Christian 2002: Mediengestaltung: Hanser, Carl, GmbH & Co.
- Gauthier, Jean-Marc 2005: Building interactive worlds in 3D. Virtual sets and pre-visualization for games, film, and the Web. Amsterdam: Elsevier/Focal Press.
- Giesen, Rolf 2000: Künstliche Welten. Tricks Special Effects und Computera-
nimation im Film von den Anfängen bis heute; Orig.-Ausg. (2000). Hamburg,
Wien: Europa-Verlag
- Giesen, Rolf 2003: Lexikon des Trick- und Animationsfilms. Von Aladdin, Aki-
ra und Sindbad bis zu Shrek, Spider-Man und South Park; Filme und Figuren,
Serien und Künstler, Studios und Technik; die große Welt der animierten Filme;
Zeichen-, Puppen-, Silhouetten- und Legetrick, Stop Motion, Mischfilme und
3D-Computeranimation. Berlin: Schwarzkopf & Schwarzkopf.
- Glassner, Andrew S. 1989: An Introduction to Ray Tracing. Academic Press.
- Goldstein, E. Bruce 2001: Pictorial perception and art. In Goldstein, E. B. (Hrsg.):
Blackwell handbook of perception, S. 344-378. Oxford, UK: Blackwell.
- Goldstein, E. Bruce 2008: Wahrnehmungspsychologie. Der Grundkurs. 7. Aufl.
Berlin, Heidelberg: Spektrum Akad. Verl.
- Harten, Ulrich 2007: Physik. Einführung für Ingenieure und Naturwissenschaft-
ler. 3., bearb. und erw. Aufl. Berlin: Springer (Springer-Lehrbuch).
- Heine, L. 1935: Anatomisch-physiologisch-pathologische Grundlagen der Stere-
oskopie. In: Naturwissenschaften, Jg. 23, H. 51, S. 855-860.
- Hoffmann, Albrecht 1990: Das Stereoskop. Geschichte der Stereoskopie. Mün-
chen: Deutsches Museum (Materialien für die betriebliche Ausbildung).
- Huygens, Christiaan 1667: Deuxième complément de la Dioptrique. In: Œuvres
Complètes, XIII/2. La Haye, Nckhoff, 1916.
- Julesz, Bela 1964: Binocular depth perception without familiarity cues. In: Sci-
ence 145, S. 356-362
- Kaplan, G. 1969: Kinetic disruption of optical texture: The perception of depth
at an edge. In: Perception and Psychophysics 6, S. 193-198
- Kepler, Johannes 1604: Ad Vitellionem paralipomena, quibus astronomiae
pars optica traditur. In: Johannes Kepler: Gesammelte Werke, Ed. Walther von
Dyck & Max Caspar, München, 1939
- Kidd, S.A. 2002: The role of qualitative research in psychological journals. In:
Psychological Methods, 7, S. 126-138.
- Knoll, John 2005: Star Wars. George Lucas und die Schöpfung seiner Welten:
Knesebeck.
- Küppers, Harald 2005: Farbenlehre. Orig.-Ausg. Köln: DuMont (DuMont-Ta-
schenbücher, 563 : DuMont-Schnellkurs).
- Kuhn, Gerhard 1999: Stereo-Fotografie und Raumbild-Projektion. Theorie und
Praxis, Geräte, Materialien. Gilching: vfv.
- Kurzer, F. 2004: The life and work of Charles Tomlinson FRS: A Career in Vic-
torian Science and Technology'. In NR 2004 Vol 58 S. 203-226
- Laube, Helene 2007: Vorwärts in die Vergangenheit. In: Financial Times Deutsch-
land, 27.03.2007. Verlag: Financial Times Deutschland GmbH & Co. KG

- Lejeune, A. 1948: Euclide et Ptolémée. Deux stades de l'optique géométrique grecque. Université de Louvain, Recueil de travaux d'histoire et de philologie, ser. 3, fasc. 31
- Lejeune, A. 1958: Les recherches de Ptolémée sur la vision binoculaire. In: Janus 1958, 47, S. 79-86
- Lippert, Herbert 1995: Anatomie. Text und Atlas; deutsche und lateinische Bezeichnungen. 6., überarb. u. erw. Aufl. München: Urban & Schwarzenberg.
- Lipton, Lenny 1982: Foundations of the Stereoscopic Cinema. A Study in Depth. Van Nostrand Reinhold Company. New York, Cincinnati, Toronto, London,
- Mach, Rüdiger & Petschek, Peter 2006: Visualisierung digitaler Gelände- und Landschaftsdaten. Berlin: Springer.
- Mäusl, Rudolf 1992: Repetitorium - Fernsehtechnik. München: Rohde & Schwarz.
- Mamassian, P. & Knill, D. & Kersten, D. 1998: The perception of cast shadows. In: Trends in Cognitive Sciences 2, S. 288-295.
- Mahintorabi, Keywan 2003: Maya. 3D-Grafik und 3D-Animation; Modellieren, Rendern, Animieren; Texturen, Materialien, Lichteffekte. 1. Aufl. Bonn: mitp-Verl. (Grafik).
- Manthey, Dirk & Fiebig, Helmut 1998a: Making of ... wie ein Film entsteht. Band 1: Idee, Produktion, Drehbuch, Storyboard & Konzept, Regie, Kamera, Schauspieler, Kostüm & Design, Marketing. Reinbek bei Hamburg: Rowohlt (Rororo, 60574 : Rororo-Sachbuch : Rororo Film + TV).
- Manthey, Dirk & Fiebig, Helmut 1998b: Making of ... wie ein Film entsteht. Band 2: Set-Team, Effekte & Tricks, Maske, Stop motion Animation, digitale Effekte, Schnitt, Ton & Musik, Synchronisation, Zukunft des Kinos. Reinbek bei Hamburg: Rowohlt (Rororo Rororo-Sachbuch, 60575).
- Marchesi, Jost J. 2002: Photokollegium. Ein Selbstlehrgang über d. technischen Grundlagen der Photographie. 5. Auflage. Schachhausen: Verl. Photographie (Band 1-6).
- McWade, John 2006: Our color wheel. In: Before & After magazine 45, S. 7. Online verfügbar unter: <http://www.bamagazine.com>, zuletzt geprüft am 10.12.2007.
- Mitchell, Rich 2004: The tragedy of 3-D cinema. In Film History: An International Journal, Vol. 16, No. 3, Pages 208-215
- Monaco, James 2007: Film verstehen. Kunst, Technik, Sprache, Geschichte und Theorie des Films und der Medien; mit einer Einführung in Multimedia. 9. Aufl., überarb. und erw. Neuausg. 2000. Reinbek bei Hamburg: Rowohlt-Taschenbuch-Verl.
- Mon-William, M. & Tresilian, J. R. 1999: Some recent studies on the extraretinal contribution to distance perception. In: Perception 28, S. 167-181.
- Mulach, Thomas & Giesen, Rolf 2002: Special visual effects. Planung und Produktion. 1. Aufl. Gerlingen: Bleicher (Produktionspraxis, 10).
- Neumann, Thomas 1966: Sozialgeschichte der Photographie. Neuwied: Luchter-

hand (Soziologische Essays).

- Newhall, Beaumont 2005: Geschichte der Photographie: München: Schirmer Mosel.
- Nischwitz, Alfred & Haberäcker, Peter 2004: Masterkurs Computergrafik und Bildverarbeitung. Alles für Studium und Praxis - Bildverarbeitungswerkzeuge. 1. Aufl. Wiesbaden: Vieweg.
- Ooi, T. L. & Wu, B. & He, Z. J. 2001: Distance determined by the angular declination below the horizon. In: Natur 414, S. 197-200.
- Oettermann, Stephan 1980: Das Panorama. Die Geschichte eines Massenmediums. Frankfurt a.M.
- Patterson, D. A. & Gibson, G. & Katz, R. H. 1988: A Case for Redundant Arrays of Inexpensive Disks (RAID). University of California, Computer Science Division. Online verfügbar unter: <http://www.cs.cmu.edu/~garth/RAIDpaper/Patterson88.pdf>. Letzter Zugriff: 12.02.2008. Auf der beigelegten CD enthalten.
- Pietsch, Werner 1957: Die Praxis der Stereo-Nahaufnahmen. Wilhelm-Knapp-Verlag, Halle/Saale
- Poynton, Charles A. 2003: Digital Video and HDTV. Algorithms And Interfaces: Morgan Kaufmann Publishing.
- Radeck, Michael 2004: Artikel: Keying am AVID DS Nitris. In: Digital Production 04:04, ACT GmbH, München, S.156-157.
- Reimers, Ulrich 1997: Digitale Fernsehtechnik. Datenkompression und Übertragung für DVB: Springer-Verlag GmbH & Co. KG.
- Rohr, Moritz von 1923: Auswahl aus der Behandlung des Horopters bei Fr. Aguilonius um 1613. In: Zeitschrift für ophthalmologische Optik 11, S. 41-59
- Rollmann, Wilhelm 1853: Zwei neue stereoskopische Methoden. In: J. C. Pogendorff (Hrsg.), Annalen der Physik (und Chemie). Halle/Leipzig. S. 186f
- Sammons, Eddie 1992: The World of 3-D Movies. A Delphi Publication.
- Schandry, Rainer 2006: Biologische Psychologie. Ein Lehrbuch. 2., überarb. Aufl. Weinheim: Beltz PVU.
- Schäffer, Arne 1996: Mensch, Körper, Krankheit. 2., überarb. und erw. Aufl., 2. korrr. Nachdr. Neckarsulm: Jungjohann.
- Schmidt, R. F. & Lang, F. & Thews†, G. 2005: Physiologie des Menschen. Mit Pathophysiologie. 29. vollständig neu bearbeitete und aktualisierte Auflage. Berlin, Heidelberg: Springer Medizin Verlag Heidelberg (Springer-11777 /Dig. Serial)).
- Schmidt, R. F. & Schaible, H.-G. & Birbaumer, N. 2006: Neuro- und Sinnesphysiologie.
- Mit 12 Tabellen. 5., neu bearb. Aufl. Heidelberg: Springer Medizin (Springer-Lehrbuch).
- Schmidt, Ulrich 2000: Professionelle Videotechnik. Analoge und digitale Grundlagen
- Signalformen Videoaufnahme Wiedergabe Speicherung Signalverarbeitung und Studio-

- technik; mit 49 Tabellen. 2., aktualisierte und erw. Aufl. Berlin, Heidelberg: Springer.
- Schönherr, Maximilian 2006: Maya 7. Die Referenz; über 1000 Befehle übersetzt, erklärt, illustriert. München: Addison-Wesley (dpi).
- Schubert, G. 1965: Visueller Cortex und Binocularsehen. In: Graefe's Archive for Clinical and Experimental Ophthalmology. Springer Berlin / Heidelberg. Volume 168, Number 5/Okttober 1965, S.431-437
- Schubert, G. 1967: Die Physiologie des Binocularsehen. In: Documenta ophthalmologica. The official journal of the International Society for Clinical Electrophysiology of Vision. Dordrecht, Dordrecht, The Hague: Springer; Kluwer Acad. Publ.; Junk. Volume 23, Number 1 / Dezember 1967 S. 1-17.
- Schwegler, Johann S. 1996: Der Mensch - Anatomie und Physiologie. Schritt für Schritt Zusammenhänge verstehen. Stuttgart: Thieme.
- Shaw, Jeffrey & Weibel, Peter 2003: Future cinema. The cinematic imaginary after film. Karlsruhe: ZKM (Electronic culture).
- Sillion, Francois X. & Puech, Claude 1994: Radiosity, Global Illumination. Morgan Kaufmann.
- Smith, Steve 2007: Perfect Color Keys: A Checklist, in Creative Cow Magazine. The Magazine for Media Professionals Working in Video, Film, Audio, Motion Graphics, Imaging & Design (May/June 2007). Publishers: Ron & Kathlyn Lindeboom. A creativecow. net Publication.
- Steinhauser A. 1889: Jahrbuch für Photographie und Reproduktionstechnik (1887-1914). Eder J. M. (Hrsg.) Halle, S.
- Stephan, Volker 2008: Bachelor-Thesis: Konzeption und Implementierung eines praxisorientierten, stereoskopischen HD-Produktionssystems. Hochschule Furtwangen University, Fakultät Digitale Medien, Wintersemester 2007/2008
- Storch, V. & Welsch, U. & Wink, M. & Arendt, D. 2007: Evolutionsbiologie. 2., vollst. überarb. und erw. Aufl. Berlin: Springer (Springer-Lehrbuch).
- Taylor J. 1750: Mechanismus des menschlichen Auges. Frankfurt
- Thissen, Frank 2000: Screen-Design-Handbuch. Effektiv informieren und kommunizieren mit Multimedia. Berlin, Heidelberg: Springer.
- Tillmanns, Urs 1991: Foto Lexikon. 1367 Fachbegriffe: Photographie.
- Tresilian, J. R. & Mon-Williams, M. & Kelly, B. 1999: Increasing confidence in vergence as a cue to distance. in: Proceedings of the Royal Society of London, 266B, S. 39-44.
- Trevor-Roper, Patrick 1989: Chevalier Taylor - Ophthalmiater Royal (1703-1772). In: Documenta Ophthalmologica, Volume 71, Number 2 / Februar 1989, S. 113-122
- Uka, T. & DeAngelis, G. C. 2003: Contribution of middle temporal area to coarse depth discrimination: Comparison of neuronal and psycho-physical sensitivity. In: Journal of Neuroscience 23, S. 3515-3530
- Wade, Nicholas J. 2005: Perception and Illusion. Historical Perspectives. Boston, MA: Springer Science + Business Media Inc.

- Walk, Richard D. & Gibson, Eleanor J. 1961: A comparative and analytical study of visual depth perception. In: Washington: American Psychological Ass. (Psychological monographs 519, S. 75
- Wheatstone, Charles 1838: Contributions to the Physiology of Vision. Part the First. On some remarkable, and hitherto unobserved, Phenomena of Binocular Vision. In: Philosophical Transactions of the Royal Society of London, Vol. 128, S. 371 - 394. Online verfügbar unter <http://www.stereoscopy.com/library/wheatstone-paper1838.html>, zuletzt geprüft am 02.04.2008
- Wheatstone, Charles 1879: The scientific papers of Sir Charles Wheatstone. Published by The Physical Society of London. London: Taylor and Francis.
- Wheatstone, Charles 2007: The Scientific Papers of Sir Charles Wheatstone: Kessinger Publishing.
- Wurtz, R. H. & Kandel, E. R. 2000: Perception of motion, depth, and form. In: Principles of neuronal science. New York: McGraw-Hill, S. 548-570

Internet

- 3-D Review 2008: The online magazine of stereoscopic product reviews and information. Online verfügbar unter: <http://www.3-dreview.com/>, zuletzt geprüft am 05.05.2008.
- 3-D Revolution Productions 2007. Firmenhomepage. Perspective on 3-D Movie Making. Online verfügbar unter: <http://www.the3drevolution.com/3danatomy.html/>, zuletzt geprüft am 20.05.2008.
- Adobe 2008a: Firmenprofil. Online verfügbar unter: <http://www.adobe.com/de/aboutadobe/>, zuletzt geprüft am 27.04.2008.
- Adobe 2008b: Features von After Effects. Online verfügbar unter: <http://www.adobe.com/de/products/aftereffects/features/>, zuletzt geprüft am 27.04.2008.
- Bahr, Achim 1993: Anaglyphen. Online verfügbar unter <http://www.3dwebsite.de/de/assets/downloads/anaglyphen.pdf>, zuletzt geprüft am 10.04.2008
- BARCO Simulation Products 2002: Stereoscopic Projection. 3D Projection Technologie. Online verfügbar unter: http://www.barco.com/projection_systems/downloads/Barco_stereoscopic_proj.pdf, zuletzt geprüft am 02.04.2008.
- Bauhaus 2008: bauhaus - archiv museum für gestaltung. Online verfügbar unter: http://www.bauhaus.de/bauhaus1919/biographien/druck_itten.htm, zuletzt geprüft am 26.02.2008.
- BFS 2008: Bundesverband Filmschnitt Editor e. V. (2008): BFS-Filmeditor. Online verfügbar unter <http://www.bfs-cutter.de/>, zuletzt aktualisiert am 30.01.2008, zuletzt geprüft am 25.02.2008.
- Biesterfeld, Lutz 2008: Förderverein für Kaiser-Panoramen e. V. Online verfügbar unter <http://www.kaiser-panorama.de>, zuletzt geprüft am 07.04.2008.
- CIE 2007: Commission Internationale de l'Eclairage. Online verfügbar unter: http://www.cie.co.at/index_ie.html, zuletzt aktualisiert am 07.01.2008, zuletzt

geprüft am 06.03.2008.

- Codex Digital 2008: Firmenhomepage. High Resolution Media Systems. Online verfügbar unter: <http://www.codexdigital.com/index.php>, zuletzt geprüft am 20.05.2008.
- Cohen, David S. 2008: James Cameron supercharges 3-D. In: Variety. Online verfügbar unter: <http://www.variety.com/VR1117983864.html>, zuletzt geprüft am 10.05.2008.
- Cohen, Karl 2000: Milestones Of The Animation Industry In The 20th Century. ANIMATION WORLD MAGAZINE. Online verfügbar unter: <http://www.awn.com/mag/issue4.10/4.10pages/cohenmilestones.php3>, zuletzt geprüft am 21.02.2008.
- Flat Black Films 2005: Rotoshop. Online verfügbar unter: <http://www.flatblackfilms.com/Rotoshop.html>, zuletzt aktualisiert am 12.10.2005, zuletzt geprüft am 22.02.2008.
- Fleischer, Max 1917: Patent US1242674. Online verfügbar unter: <http://www.freepatentsonline.com/1242674.pdf>, zuletzt geprüft am 04.03.2008.
- FMX 2008: Homepage der fmx/08: 13th International Conference on Animation, Effects, Games and Digital Media. May 05-08, 2008 in Stuttgart. Online verfügbar unter: www.fmx.de, zuletzt geprüft am 11.05.2008.
- Garber, Daniel 2003: Routledge Encyclopedia of Philosophy Online: Descartes, René (1596-1650). Online verfügbar unter: <http://www.rep.routledge.com/article/DA026>, zuletzt geprüft am 01.04.2008.
- Gebhard, C. & Voigt-Müller, G. 2008a: Chancen fürs Kino: 3D. Online verfügbar unter: http://www.film-tv-video.de/newsdetail.html?&no_cache=1&uid=37252&L=, zuletzt geprüft am 05.05.2008.
- Gebhard, C. & Voigt-Müller, G. 2008b: NAB2008-Trend: Stereo-3D. Online verfügbar unter: http://www.film-tv-video.de/newsdetail.html?&no_cache=1&uid=37220&L=, zuletzt geprüft am 05.05.2008.
- Gebhard, C. & Voigt-Müller, G. 2008c: NAB2008-Trend: Überall Flash-Speicher. Online verfügbar unter: http://www.film-tv-video.de/newsdetail.html?&no_cache=1&uid=37219&L=, zuletzt geprüft am 20.05.2008.
- Gebhard, C. & Voigt-Müller, G. 2008d: NAB2008: Portabler 4K-Recorder von Codex. Online verfügbar unter: http://www.film-tv-video.de/newsdetail.html?&no_cache=1&uid=37276&L=, zuletzt geprüft am 20.05.2008.
- Hagler, Jürgen 2007: Jürgen Hagler: Digital Media for Artists. Geschichte der Animation - Die wichtigsten Entwicklungsschritte ab 1900. Kunstuniversität Linz. Online verfügbar unter: <http://www.dma.ufg.ac.at/app/link/Grundlagen:3D-Grafik/module/14100.jsessionid=91706A9D2B0420CD49503394E6A904DA>, zuletzt geprüft am 21.02.2008.
- HDV 2007: HDV Format Main Specifications (Ver. 1.0). Online verfügbar auf der HDV Information Web Site unter: <http://www.hdv-info.org/HDVSpecifications.pdf>. Letzter Zugriff: 13.02.2008. Auf der beigelegten CD enthalten.

- Jachèl, D. 2003: Stereodisplays & autostereoskopische Displays. Lehrmaterial zum Vortragsseminar Visuelle Wahrnehmung und 3D-Displays am Lehrstuhl für Visual Computing und Computergraphik der Universität Rostock. Online verfügbar unter: www.icg.informatik.uni-rostock.de/Lehre/HCG/scripte0304/04_3DDisplays.pdf, zuletzt geprüft am 28.02.2008.
- Kirsner, Scott 2008: Tech Impact Report. James Cameron, Filmmaker. In: Variety. Online verfügbar unter: <http://www.variety.com/article/VR1117978512.html?categoryid=2438&cs=1>, zuletzt geprüft am 10.05.2008.
- Küppers, Harald 2001: Küppers-Farbenlehre: Biographie H.K. Online verfügbar unter: <http://www.ipsi.fraunhofer.de/Kueppersfarbe/de/kueppers.html>, zuletzt aktualisiert am 2001, zuletzt geprüft am 26.02.2008.
- Kunz, Thomas B. 2007: TBK: Anaglyphen-Stereoskopien & Kaiserpanorama. Online verfügbar unter: <http://www.tbk.de/3dstereo/>, zuletzt geprüft am 20.02.2008.
- Lataille, Roland 2008: Cinerama. Online verfügbar unter: <http://cinerama.topcities.com/>, zuletzt geprüft am 06.05.2008.
- Lipton, Lenny 2007: Vertical Surround Parallax Correction. Online verfügbar unter: http://www.reald-corporate.com/_resources/vertical.pdf, zuletzt geprüft am 12.05.2008.
- Markstein, Don 2008a: Max Fleischer Studio - Don Markstein's Toonopedia. Online verfügbar unter: <http://www.toonopedia.com/fleischr.htm>, zuletzt geprüft am 21.02.2008.
- Markstein, Don 2008b: Glossary - Don Markstein's Toonopedia. Online verfügbar unter: <http://www.toonopedia.com/glossary.htm#rotoscope>, zuletzt geprüft am 21.02.2008.
- Matteworld 2008a: Effects For Film. Online verfügbar unter: <http://www.matteworld.com/film/index.html>, zuletzt geprüft am 20.02.2008.
- Matteworld 2008b: Company Info. Online verfügbar unter: <http://www.matteworld.com/info/>, zuletzt geprüft am 20.02.2008.
- Müller, Christine Xuân 2006: 5D-Kino: Das große Krabbeln. Online verfügbar unter: <http://www.spiegel.de/kultur/kino/0,1518,396407,00.html>, zuletzt geprüft am 20.05.2008.
- Näther, Markus 2008: 3D: Kino zum Anfassen. Online verfügbar unter: http://www.film-tv-video.de/newsdetail.html?&no_cache=1&uid=37097&L=, zuletzt geprüft am 05.05.2008.
- Pastoor, Siegmund & Wöpking, Matthias 2002: 3-D Displays: A review of current technologies. Online verfügbar unter: [http://www.dgp.toronto.edu/~gf/Research/Volumetric %20UI/3-D%20Displays%20A%20review%20of%20current%20technologies.htm](http://www.dgp.toronto.edu/~gf/Research/Volumetric%20UI/3-D%20Displays%20A%20review%20of%20current%20technologies.htm), zuletzt geprüft am 10.04.2008.
- Planck, Max 2008: Max Planck Institute for the History of Science, Berlin: The Virtual

- Laboratory Essays and Resources on the Experimentalization of Life VL People. Online verfügbar unter <http://vlp.mpiwg-berlin.mpg.de/people/data?id=per90>, zuletzt geprüft am 26.02.2008.
- Poch, Bernd 1999: Das Kaiserpanorama. Das Medium, seine Vorgänger und seine Verbreitung in Nordwestdeutschland. Online verfügbar unter: <http://www.uni-oldenburg.de/kunst/mediengeschichte/kaiserpanorama/emden/emden.htm>
- Mediaprofis 2008: HD - Die neue Dimension des Film(en)s. Online verfügbar unter: <http://www.mediaprofis.net/content/view/132/49/>, zuletzt geprüft am 13.02.2008.
- PresseBox 2003: DVC stellt Baytechs CineRAM erstmalig auf dem Open-House Event vom 8. - 9. Juli 2003 vor. Huber Verlag für Neue Medien GmbH. Online verfügbar unter: <http://www.pressebox.de/pressemeldungen/dvc-digitalvideo-computing-gmbh/boxid-8415.html>, zuletzt geprüft am 18.02.2008.
- Quantel 2008a: Firmenhomepage. Online verfügbar unter: <http://www.quantel.com/>, zuletzt geprüft am 05.05.2008.
- Quantel2008b:Stereoscopic3D.Addanewdimensiontoyourbusiness.Onlineverfügbar unter: [http://www.quantel.com/resource.nsf/Files/Stereoscopic_3D_nab08/\\$FILE/Stereoscopic_3D_nab08.pdf](http://www.quantel.com/resource.nsf/Files/Stereoscopic_3D_nab08/$FILE/Stereoscopic_3D_nab08.pdf), zuletzt geprüft am 05.05.2008.
- Radeck, Michael 2003: Radeck, Michael: Biografie. Online verfügbar unter: <http://www.michael-radeck.de/biografi.html>. Zuletzt aktualisiert am 09. Juli 2003, zuletzt geprüft am 25.02.2008
- Sabiston, Bob 2001: Bob Sabiston: G4 - Feature - Rotoshop Animation. Online verfügbar unter: http://www.g4tv.com/techtvvault/features/23423/Rotoshop_Animation_pg2.html, zuletzt aktualisiert am 16. October 2001, zuletzt geprüft am 22.02.2008.
- SabuCat Productions 2006: 2006 World 3-D Film Festival. Online verfügbar unter: <http://www.3dfilmfest.com/>, zuletzt geprüft am 05.05.2008.
- Screen Digest 2008: Firmenhomepage der Analysefirma. Online verfügbar unter: <http://www.screendigest.com/>, zuletzt geprüft am 05.05.2008.
- Senf, Erhard 2007: Das Kaiser-Panorama. Was ist das? Online verfügbar unter: <http://www.tbk.de/3dstereo/kpvskp.htm>, zuletzt geprüft am 02.04.2008
- Storz Eventfabrik GmbH 2005: Prime Cinema 5D. Online verfügbar unter: <http://www.5dkino.at/index.php>, zuletzt geprüft am 02.06.2008
- Texas Instruments 2008: How DLP technology works. Online verfügbar unter: <http://www.dlp.com/tech/what.aspx>, zuletzt geprüft am 10.05.2008
- The Foundry Visionmongers Ltd. 2008: Homepage der Entwicklerfirma für hochspezialisierte Visual Effects Software. Online verfügbar unter: <http://www.thefoundry.co.uk/index.aspx>, zuletzt geprüft am 24.04.2008
- Tridality Display Solutions GmbH 2008: Firmenhomepage. Online verfügbar unter: <http://www.tridality.de>, zuletzt geprüft am 05.05.2008
- Turner, Henry 2008: Hollywood rediscovers 3-D technology. In: Variety. Online verfügbar unter: <http://www.variety.com/article/VR1117973878.html?categoryid=2731&cs=1&nid=2562>, zuletzt geprüft am 10.05.2008.

- Wacom Co. Ltd. 2008: Hersteller von Grafiktablets mit druckempfindlichen Stift ohne Batterie und Kabel. Die Firmenhomepage ist online verfügbar unter: <http://www.wacom-europe.com>, zuletzt geprüft am 25.04.2008
- Warner 2008: Warner Independent Pictures' A Scanner Darkly -- The Official Film site. Online verfügbar unter <http://wip.warnerbros.com/ascannerdarkly/>, zuletzt geprüft am 22.02.2008
- Wettmann, Hartmut; Berlin 2007: Stereoskopiegeschichte. Online verfügbar unter [http:// fotoplatz.stereographie.de/stereoskopie/](http://fotoplatz.stereographie.de/stereoskopie/), zuletzt aktualisiert am 06.10.2007, zuletzt geprüft am 06.04.2008.
- Wheatstone, Charles 1838: The Library: Wheatstone Paper 1838. Contributions to the Physiology of Vision. Part the First. On some remarkable, and hitherto unobserved, Phenomena of Binocular Vision. Online verfügbar unter: <http://www.stereoscopy.com/library/wheatstone-paper1838.html>, zuletzt geprüft am 02.04.2008
- Wikipedia2007: Artikel Bob Sabiston. In Wikipedia, The Free Encyclopedia. Bearbeitungsstand: 30. November 2007. Retrieved 13:00, February 22, 2008, from http://en.wikipedia.org/w/index.php?title=Bob_Sabiston&oldid=174867309
- Wikipedia 2008a: Artikel RAID. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 6. Februar 2008, 15:19 UTC. URL: <http://de.wikipedia.org/w/index.php?title=RAID&oldid=42169091> (Abgerufen: 13. Februar 2008, 12:43 UTC)
- Wikipedia 2008b: Artikel HDCAM. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 16. Februar 2008, 17:57 UTC. URL: <http://de.wikipedia.org/w/index.php?title=HDCAM&oldid=42596265> (Abgerufen: 18. Februar 2008, 13:37 UTC)
- Wikipedia 2008c: Artikel Max Fleischer. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 2. Februar 2008, 09:41 UTC. URL: http://de.wikipedia.org/w/index.php?title=Max_Fleischer&oldid=41954368 (Abgerufen: 21. Februar 2008, 14:30 UTC)
- Wikipedia 2008d: Artikel Rotoskopie. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 24. Januar 2008, 07:15 UTC. URL: <http://de.wikipedia.org/w/index.php?title=Rotoskopie&oldid=41570494> (Abgerufen: 21. Februar 2008, 15:06 UTC)
- Wikipedia 2008e: Artikel A Scanner Darkly - Der dunkle Schirm. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 9. Februar 2008, 11:13 UTC. URL: [http:// de.wikipedia.org/w/index.php?title=A_Scanner_Darkly_%E2%80%93_Der_dunkle_Schirm&oldid=42285496](http://de.wikipedia.org/w/index.php?title=A_Scanner_Darkly_%E2%80%93_Der_dunkle_Schirm&oldid=42285496) (Abgerufen: 22. Februar 2008, 12:57 UTC)
- Wikipedia 2008f: Artikel Rotoshop. In Wikipedia, The Free Encyclopedia. Bearbeitungsstand: 22. Januar 2008. URL: <http://en.wikipedia.org/w/index.php?title=Rotoshop&oldid=186115694> (Abgerufen: 22. Februar 2008, 13:05 UTC)
- Wikipedia 2008g: Artikel Laurens Hammond. In Wikipedia, The Free Ency-

- clopedia. Bearbeitungsstand: March 22, 2008. Online verfügbar unter: http://en.wikipedia.org/w/index.php?title=Laurens_Hammond&oldid=200170853, zuletzt geprüft am 05.05.2008.
- Wikipedia 2008h: Artikel Bwana Devil. Wikipedia, The Free Encyclopedia. Bearbeitungsstand: April 29, 2008, 14:22 UTC. Available at: http://en.wikipedia.org/w/index.php?title=Bwana_Devil&oldid=208999471, zuletzt geprüft am 06.05.2008.
 - Wimmer, Peter 2008: Homepage von 3dtv.at, Entwicklerfirma des Stereoscopic Players. Online verfügbar unter: <http://www.3dtv.at>, zuletzt geprüft am 06.04.2008.
 - Wittkowsky, Wilfried 2003: Der 3D-Film. Spielfilme. Online verfügbar unter: <http://www.wittkowsky.net/3d-film/spiel.htm>, zuletzt geprüft am 06.05.2008.
 - Woods, Andrew 2008: The virtual library of the Stereoscopic Displays and Applications conference website. Online verfügbar unter: <http://www.stereoscopic.org/>, zuletzt geprüft am 06.05.2008
 - Woods, Andrew J. & Rourke, Tegan 2004: Ghosting in Anaglyphic Stereoscopic Images. In: Stereoscopic Displays and Virtual Reality Systems XI, Proceedings of SPIE-IS&T Electronic Imaging, SPIE Vol. 5291, San Jose, California. Online verfügbar unter: <http://www.cmst.curtin.edu.au/publicat/2004-08.pdf>, zuletzt geprüft am 05.04.2008
 - Yahoo Groups 2008: Group 3D Video. Online verfügbar unter: <http://de.groups.yahoo.com/group/3d-video/>, zuletzt geprüft am 15.05.2008 3. Weiterführend Internetlinks
 - Adam, Hans Christian 2007: PhotoLit. An international databank which aims to list photographic literature published since 1839. Online verfügbar unter: <http://www.schaden.com/photolit/index.html>, zuletzt geprüft am 07.04.2008.
 - Bahr, Achim 2008: Stereoskopie und Visualisierung. Online verfügbar unter: <http://www.3dwebsite.de>, zuletzt geprüft am 10.04.2008
 - Bungert, Christoph 2005: The independent 3D and Virtual Reality resource. Online verfügbar unter: <http://www.stereo3d.com/>, zuletzt geprüft am 05.05.2008
 - EEFX 2006: Firmenhomepage. Chroma Key products for the visual effects industry. Online verfügbar unter: <http://www.eefx.com/>, zuletzt geprüft am 01.05.2008
 - Gavrilov, Kirill 2008: Entwicklerseite von sView, einem stereoskopischen Bildbetrachter. Online verfügbar unter: <http://www.sview.ru/en/sview/info>, zuletzt geprüft am 29.04.2008
 - Herbig, Gerhard P. 2005: Technische Grundlagen zur Stereofotografie. Online verfügbar unter: <http://www.herbig-3d.de>, zuletzt geprüft am 06.04.2008
 - Kiss, Oliver 2006: Tutorial zum Erstellen von 3D-Bildern. Online verfügbar unter: http://www.hyperstream.de/index.php?option=com_content&task=view&id=6&Itemid=5/ zuletzt geprüft am 14.04.2008

- Köhler, André 2008: Die Erzeugung des Wheatstoneschen Stereoeffektes und Stereofotografie. Online verfügbar unter: <http://www.koehlerandre.hosting-le.de/ste/ste02.html>, zuletzt geprüft am 02.04.2008
- KUK Filmproduktion 2006: 3D- Film Produktion. Online verfügbar unter: <http://www.kuk-film.de/html/start.html>, zuletzt geprüft am 07.05.2008.
- Lottenburger, Helmut 2007: Private Seite zum Thema Chromakeying. Online verfügbar unter: <http://www.chromakeying.com/>, zuletzt geprüft am 15.05.2008.
- Markus Schaden 2008: Buchhandlung für fotografische Fachpublikationen. Online verfügbar unter: <http://www.schaden.com/index.php>, zuletzt geprüft am 07.04.2008.
- NASA 2008: STEREO (Solar TERrestrial RELations Observatory). Online verfügbar unter: http://www.nasa.gov/mission_pages/stereo/main/index.html, zuletzt geprüft am 10.05.2008
- Planding, Wolfgang 2002: Stereoskopische Mondbilder. Online verfügbar unter: <http://www.mondatlas.de/stereoskopie/stereoskopie.html>, zuletzt geprüft am 06.05.2008
- Real D 2007: Firmenhomepage des Anbieters von Stereo-3D-Kinoprojektoren. Online verfügbar unter: <http://www.reald-corporate.com/>, zuletzt geprüft am 09.05.2008
- Schneider, A. & Brugger, J. & Beyer, P. 2006: Analyse von 3D Stereo Rendering und Motion-Tracking. Online verfügbar unter: <http://projects.cynapses.org/StereoRender/paper.pdf>, zuletzt geprüft am 09.04.2008
- Schulte, Dieter 2007: Lexikon der Stereoskopie. Online verfügbar unter: http://www.stereoskopie.com/Lexikon_der_Stereoskopie/body_lexikon_der_stereoskopie.html, zuletzt aktualisiert am 07.08.2007, zuletzt geprüft am 06.04.2008.
- Schulte, Dieter 2008: Virtuelles Museum für Stereofotografie und Stereoskopie. Größte deutsche Website für Stereoskopie mit über 900 Seiten und ca. 6000 Abbildungen. Online verfügbar unter: <http://www.stereoskopie.com/>, zuletzt geprüft am 10.04.2008.
- Shelly, Dan 2006: Homepage mit über 600 Links zur Stereoskopie. Online verfügbar unter: <http://www.dddesign.com/3dbydan/default.htm>, zuletzt geprüft am 09.05.2008
- Starosta, Boris 2005: Stereoscopic Art & 3-D Photography. Online verfügbar unter: <http://www.starosta.com>, zuletzt geprüft am 06.04.2008
- Thies, Heiko 2008: Amateurfilme und visuelle Effekte. Online verfügbar unter: <http://www.hethfilms.de/?show=videos>, zuletzt geprüft am 16.05.2008
- Wilson, Terry 2008: Stereo Artist. Online verfügbar unter: <http://www.terry-fic3d.com>, zuletzt geprüft am 06.04.2008.
- Wimmer, Peter 2002: Technik stereoskopischer Videos. Online verfügbar unter: <http://pwimmer.gmxhome.de/LoadFrames.html?/3D/3DVideoTechnique.html>, zuletzt geprüft am 20.04.2008
- Wirths, Karsten 2002: Stereofotografie gestern und heute. Online verfügbar unter: <http://www.stereoblick.de/>, zuletzt geprüft am 05.05.2008

- Wittkowsky, Wilfried: Der 3D-Film. Online verfügbar unter: <http://www.wittkowsky.net/3d-film/index.htm>, zuletzt geprüft am 05.05.2008
- Zang, Christoph 2005: Facharbeit: Stereoskopie. Online verfügbar unter: <http://www.christoph-zang.de/Facharbeit.32.0.html>, zuletzt geprüft am 02.04.2008
- Ziegler, Christian 1999: Facharbeit aus dem Fach Physik: Stereofotografie. Online verfügbar unter: <http://www.ckg-wuerzburg.de/schule/faecher/physik/facharb/zieglchr/stereo.php>, zuletzt geprüft am 02.04.2008

Teil 5

Untersuchungen zur automatischen Generierung von Tiefenkarten in der S3D-Produktion

Eric Schumacher

Abbildungsverzeichnis

Abbildung 1: Tiefenkarten	286
Abbildung 2: Depth Key	287
Abbildung 3: Texturverläufe	292
Abbildung 4: Problembereiche der Korrespondenzanalyse.....	295
Abbildung 5: Reihenfolgebedingung	299
Abbildung 6: Normierte Kreuzkorrelation	301
Abbildung 7: Segmentseiten	309
Abbildung 8: Segmentkombinationen	310
Abbildung 9: Verwendetes Stereorig	315
Abbildung 10: Tiefenkarte aus Stereotracer	318
Abbildung 11: Side-by-Side-Anordnung	319
Abbildung 12: Tiefenkarte aus Side-by-Side	319
Abbildung 13: Stereoverarbeitungskette.....	320
Abbildung 14: Tiefenkarte Keller 01	321
Abbildung 15: Tiefenkarte Schnittraum 01	321
Abbildung 16: Tiefenkarte Schnittraum 02_Fahrt	322
Abbildung 17: Tiefenkarte Schulhof 01	322
Abbildung 18: Tiefenkarte Wand 01	323
Abbildung 19: Tiefenkarte Wand 02_1	323
Abbildung 20: Tiefenkarte Wand 02_2	324
Abbildung 21: Tiefenkarte Wand 02_3	324
Abbildung 22: Tiefenkarte Wand 03_1	325
Abbildung 23: Tiefenkarte Wand 03_2	325
Abbildung 24: Tiefenkarte Wand 03_3	326
Abbildung 25: TK Analyse Keller 01	326
Abbildung 26: TK Analyse Schnittraum 01	327
Abbildung 27: TK Analyse Schnittraum 02_Fahrt	328
Abbildung 28: TK Analyse Schulhof 01	329
Abbildung 29: TK Analyse Wand 01	330
Abbildung 30: TK Analyse Wand 02_1	331
Abbildung 31: TK Analyse Wand 02_1	332
Abbildung 32: TK Analyse Wand 02_3	333
Abbildung 33: TK Analyse Wand 03_1	334
Abbildung 34: TK Analyse Wand 03_2	335
Abbildung 35: TK Analyse Wand 03_3	336

1 Einführung

Der stereoskopische Spielfilm stellte seit seiner Einführung um 1950 eher einen Randbereich der Filmindustrie dar. Aktuell jedoch erfährt dieser Bereich wieder einen starken Schub, immer mehr Produktionen werden stereoskopisch umgesetzt. Während sich die technische Stereoskopie in der Filmindustrie momentan aus ihrem Nischendasein zu lösen scheint, wird sie in anderen Industriezweigen z.T. seit vielen Jahren genutzt. Es finden sich stereoskopische Verfahren z.B. in der Landvermessung oder in der industriellen Bildverarbeitung¹. Im Bereich der Computer Vision werden Stereosysteme u.a. genutzt, um autonomen Robotern maschinelles Sehen z.B. zur Navigation zu ermöglichen². Dabei wird oft ein Prozess durchgeführt, bei welchem aus den Stereobildpaaren eine sog. Tiefenkarte (depth map, z-map) generiert wird. Hierbei handelt es sich um ein Graustufenbild, bei dem über die Grauwerte die realen Tiefenwerte der abgebildeten Szene dargestellt werden³. In der Praxis existieren zwei Varianten, bei der ersten Variante repräsentieren helle Bereiche eine große Distanz zur Kamera und dunkle Bereiche eine kleine Distanz. Bei der zweiten Variante verhält es sich umgekehrt: Je heller ein Bereich ist, desto geringer ist die jeweilige Distanz zur Kamera, dunkle Bereiche liegen dabei weiter entfernt⁴. In diesem Artikel wird die zweite Variante herangezogen. Tiefenkarten werden auch in der 3D-Computergrafik verwendet, hier werden sie über einen sog. Z-Buffer-Algorithmus erzeugt⁵. Ihre Funktion besteht darin, virtuelle Kameras mit Informationen über die räumliche Beschaffenheit künstlich erzeugter Szenen und Objekte (computer generated image, kurz: CGI) zu versorgen.

1 Vgl. Wöhler 2009, S. 243ff.

2 Vgl. Vincze et al. 2008.

3 Vgl. Smirnov et al. 2010.

4 Vgl. Mach; Petschek 2006, S. 286.

5 Vgl. Bender; Brill 2006, S. 67.

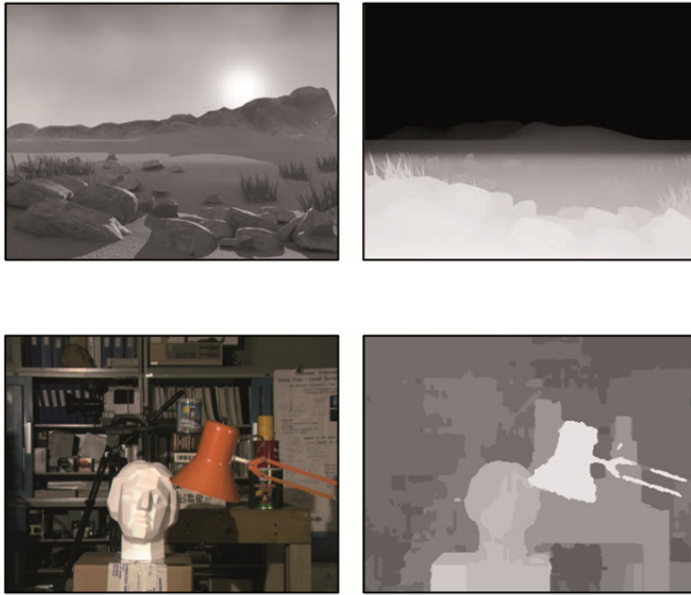


Abbildung 1: Tiefenkarten

1.1 Einsatzbereiche von Tiefenkarten

Die einzelnen Verfahren zur Generierung und Weiterverarbeitung von Tiefenkarten sind bereits ausführlich erforscht und dokumentiert, problematisch gestaltet sich oftmals die praxisorientierte Adaption der Verfahren auf die technischen Gegebenheiten der Bildaufnahmesysteme für Film- und Videoproduktion. Daher werden die bisher existierenden Herangehensweisen nur zum Teil bereits in der Produktionspraxis eingesetzt. Nachfolgend finden sich einige mögliche Anwendungsszenarien für Tiefenkarten im Film- und Videobereich.

- **Keying:** Dieser Begriff bezeichnet eine Tricktechnik in der Filmproduktion, bei der Bildelemente von ihrem Hintergrund freigestellt werden. Die Bildpunkte des Hintergrundes werden dabei mit technischen Hilfsmitteln nach verschiedenen Kriterien (z.B. Farbton oder Helligkeitswerte) separat erfasst und in ihrer Transparenz beeinflusst. Anstelle des Hintergrundes kann dann ein anderes beliebiges Motiv eingefügt werden. Keyingverfahren können während des Livebetriebs durch entsprechende Bildmischer oder in der Nachbearbeitung eingesetzt werden. Liegt von einer Szene eine Tiefenkarte vor, so kann diese zum Keying eingesetzt werden, indem die Bildpunkte des Hintergrundes abhängig von ihrem Tiefenwert transparent gemacht werden (Depth Key).

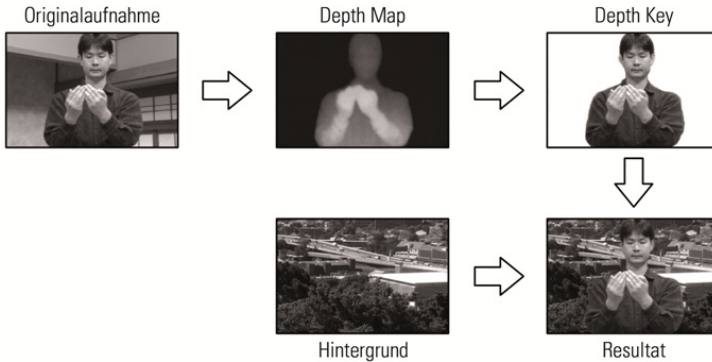


Abbildung 2: Depth Key

- **Matchmoving:** Hier handelt es sich um eine Arbeitsphase in der Postproduktionskette von Filmproduktionen. Sie ist dann notwendig, wenn CGI-Elemente für den Zuschauer glaubhaft in das real aufgenommene Material eingefügt werden sollen. Dabei werden die Bewegungen und Parameter(änderungen) der realen Kamera analysiert, gespeichert und anschließend auf die virtuelle Kamera eines 3D-Programms übertragen⁶. So stellt die virtuelle Kamera die CGI-Objekte, entsprechend der Perspektive der realen Kamera dar. Oft fällt es dabei in den Aufgabenbereich des Matchmovers, die physikalische Geometrie der realen Szene zumindest in groben Zügen zu erfassen und mit Hilfe eines 3D-Programms zu rekonstruieren⁷. Unter anderem benötigt der Matchmover hierzu die Entfernung zwischen der Kamera und den Szenenobjekten, was i.d.R. bedeutet, dass bereits am Set verschiedene Abmessungen vorgenommen werden müssen. Sind die Distanzen innerhalb der Szene jedoch mit der Hilfe von Tiefenkarten ermittelbar, so lassen sich diese Vermessungsarbeiten vermeiden. Soll nun im Rahmen einer stereoskopischen Filmproduktion Matchmoving durchgeführt werden, kann die Tiefeninformation, welche in den Kamerabildepaaren enthalten ist, mit geeigneten Verfahren extrahiert und zur Bestimmung der Szenentiefe und somit zur Rekonstruktion der Szenengeometrie genutzt werden.

- **Distribution von Stereoinhalten (2D-plus-depth):** Am Einsatz von Tiefenkarten in Zusammenhang mit der Distribution von stereoskopischen Inhalten wird derzeit ebenfalls weiter geforscht. Das entsprechende Formatkonzept sieht vor, dass bei der Verteilung z.B. über Funk- oder Kabelnetze nur die Bilder für eine der beiden Stereoansichten gemeinsam mit einer zugehörigen Tiefen-

6 Vgl. Dobbert 2005, S. 2.

7 Vgl. Dobbert 2005, S. 134.

karte übertragen werden⁸. Das Endgerät, bzw. eine vorgeschaltete technische Komponente errechnet aus diesen Informationen die fehlende zweite Ansicht (depth-image based rendering, kurz: DIBR), so dass beide Ansichten als Stereobildpaar wiedergegeben werden können⁹. Da sich eine Tiefenkarte lediglich aus Grauwert-Informationen zusammensetzt, benötigt sie bei der Übertragung weniger Bandbreite als ein Farbbild¹⁰. Ein weiterer Vorteil besteht darin, dass zusätzlich Zwischenansichten für Multiview-Anwendungen interpoliert werden können¹¹. Es ist bei diesem Verfahren darüber hinaus sofort eine Abwärtskompatibilität sichergestellt für diejenigen 2D-Endgeräte, die keine Stereoansicht darstellen können. Diese können trotzdem mit besagtem Format beliefert werden und nur die reguläre monoskopische Ansicht extrahieren.

- Depth Maps zur Bildpaarkorrektur: In der Einführung zu den Grundlagen der Stereoskopie wird darauf eingegangen, wie stereoskopische Aufnahmen von einer Vielzahl an Fehlerquellen beeinträchtigt werden können. Verschiedene Unternehmen arbeiten derzeit an Lösungen, die derartige Fehler entweder bereits bei der Aufnahme oder bei der späteren Nachbearbeitung automatisch korrigieren. Die Abweichungen zwischen den zwei Ansichten eines Stereobildpaares werden dabei mittels einer sog. Disparitätskarte erfasst. Diese wird erzeugt, indem die Positionen zweier Bildpunkte, welche die Abbildungen desselben Objektpunktes sind, miteinander verglichen werden¹². Zwischen Disparität und Raumtiefe besteht eine Beziehung, auf welche in den folgenden Abschnitten noch genauer eingegangen werden wird. Durch diesen Zusammenhang beinhaltet eine Disparitätskarte ebenfalls eine Repräsentation der Raumtiefe. Erfolgen die Korrekturen der Bildfehler bereits automatisch in Echtzeit bei der Aufnahme, muss das Aufnahmesystem entsprechend leistungsfähig sein, da für jedes Einzelbild die Erstellung der Tiefenkarte sowie die Korrekturmechanismen innerhalb eines extrem kurzen Zeitraums durchgeführt werden müssen¹³.

8 Vgl. SMPTE (Hg.) 2009, S. 37f.

9 Vgl. Smirnov et al. 2010, S. 16.

10 Vgl. Wikipedia: 2D-plus-depth.

11 Vgl. Smolic et al. 2008.

12 Vgl. Hallpike 2009.

13 Vgl. Zilly; Eisert; Kauff 2009.

2 Tiefenbilder

2.1 Prinzipien der Tiefenbildaufnahme

Die Erstellung von Tiefenbildern für eine Szene kann gleichgesetzt werden mit der Erfassung der dreidimensionalen Struktur der in der Szene sichtbaren Oberflächen. Diese Tiefenbilder unterscheiden sich von den sog. Volumenbildern, bei denen es um die vollständige dreidimensionale Rekonstruktion eines Szenenobjekts geht, d.h. es werden auch Strukturen erfasst, die sich hinter den sichtbaren Oberflächen befinden¹⁴. Bei der Tiefenbildgewinnung kann jeweils zwischen aktiven und passiven Verfahren unterschieden werden. Aktive Verfahren sind i.d.R. aufwändiger, da sie zusätzlich zur Kamera weitere Komponenten benötigen. Dieser Mehraufwand führt aber auch zu einer Tieferfassung, die genauer ist als bei den passiven Verfahren. Bei den passiven Verfahren handelt es sich um unterschiedliche interne Bildverarbeitungsprozesse, die eine Erfassung der Tiefe direkt aus dem Inhalt der aufgenommenen Sequenzen ermöglichen.

2.1.1 Aktive Verfahren

2.1.1.1 Tiefe aus Laufzeit

Ein Sendegerät sendet in Richtung der Szene ein Signal aus, welches sich mit einer bekannten Geschwindigkeit fortpflanzt. Die Szenenobjekte reflektieren das Signal und senden es dadurch zu einem Empfangsgerät zurück. Dort wird die Zeit, in der das Signal unterwegs war, gemessen und ausgewertet. In der Regel befinden sich Sender und Empfänger auf der Höhe der Kamera, die Laufzeit des Signals ist dann proportional zur Summe der Entfernungen zwischen Sender und Objekt und Objekt und Empfänger¹⁵. Als Signalgeber werden beispielsweise Ultraschallgeber oder Infrarotlichtquellen genutzt und als Empfänger Mikrofonarrays oder lichtempfindliche Sensoren. Ein Beispiel für ein optisch basiertes System ist die HDTV Axi-Vision Camera¹⁶, die als Signalquelle LED-Arrays (light emitting diode) nutzt. Als Empfänger dient eine CCD-Kamera mit einem Hochgeschwindigkeits-Shutter. Die LEDs senden Strahlen im Nah-Infrarotbereich (850 nm) aus, d.h. sie sind im eigentlichen Kamerabild nicht sichtbar. Allgemein werden derartige Systeme unter dem Begriff TOF-Kamera zusammengefasst (time of flight, kurz: TOF).

2.1.1.2 Tiefe aus Phase

Diese Verfahren werden zusammengefasst unter dem Fachbegriff Interferometrie und nutzen ebenfalls Sende- und Empfangssysteme. Im Gegensatz zur beschriebenen Tiefermittlung aus Laufzeit wird hier nicht nur die Amplitude (d.h. Energie) der

¹⁴ Vgl. Jähne 2005, S. 227.

¹⁵ Vgl. Jähne 2005, S. 229.

¹⁶ Vgl. Kawakita et al. 2004.

Strahlung gemessen, sondern auch ihre Phasenlage¹⁷. Hierzu wird der Effekt der Interferenz genutzt. Dieser ist dann gegeben, wenn sich zwei oder mehr Wellen gemäß dem sog. Superpositionsprinzip überlagern. Dieses wiederum besagt, dass sich zwei unabhängige elektromagnetische Felder linear zu einem Superpositionsfeld überlagern¹⁸. Löschen sich die Wellen gegenseitig aus, wird dies als destruktive Interferenz bezeichnet, bei einer Verstärkung der Amplituden handelt es sich um konstruktive Interferenz. Zur Tiefenmessung werden zwei oder mehr kohärente Wellen genutzt, d.h. es herrscht eine systematische Beziehung zwischen ihren Phasen¹⁹. Die von der Szene reflektierten Wellen erzeugen bei der Empfangseinheit Intensitätsvariationen (konstruktive und destruktive Interferenzen), wodurch die Messung der Phasenlage und damit der Tiefe ermöglicht wird. Zur Erzeugung kohärenter Lichtwellen werden oft Laser eingesetzt.

2.1.1.3 Tiefe aus Kohärenz

Bei der Überlagerung von zwei oder mehr Wellen wird die Gesamtheit der Bereiche mit destruktiver und konstruktiver Interferenz als Interferenzmuster bezeichnet²⁰. Die Fähigkeit zur Bildung eines stabilen Interferenzmusters hängt von der Kohärenzzeit der Wellen ab. Diese gibt an, wie stark die Wellen maximal gegeneinander verschoben sein dürfen, bevor das Muster instabil wird²¹. Aus der Kohärenzzeit leitet sich die Kohärenzlänge ab; diese entspricht der optischen Strecke, die das Licht während der Kohärenzzeit zurücklegt. Dadurch, dass Kohärenzlängen nur wenige Wellenlängen kurz sein können, schließen Tiefenmessungen aus Kohärenz die Lücke zwischen Messungen aus Phasenlage und Messungen aus Laufzeit²². Ein Beispiel für die Entfernungsmessung mittels Kohärenz sind Weißlichtinterferometer²³.

2.1.2 Passive Verfahren

2.1.2.1 Tiefe aus Schattierung

Bei den schattierungsbasierten Verfahren wird eine Beziehung zwischen den Variationen der Grauwertverteilung im Bild, der Oberflächeneigenschaften der Szenenobjekte und der Beleuchtung hergestellt. Wird von einer bekannten konstanten Reflexion und einer bekannten Beleuchtung ausgegangen, so können aus den Bildintensitäten die Normalenvektoren der Oberflächen rekonstruiert und daraus die relativen Tiefenwerte für alle Pixel berechnet werden. Hierzu kann das Konzept einer sog. Lambertschen Oberfläche genutzt werden²⁴. Es handelt sich hierbei um eine perfekt diffus reflektierende Fläche, d.h. einfallende Lichtstrahlen werden gleichmäßig in alle

¹⁷ Vgl. Meschede 2008, S. 179.

¹⁸ Vgl. Meschede 2008, S. 41.

¹⁹ Vgl. Jähne 2005, S. 181.

²⁰ Vgl. Wikipedia: Interferenz (Physik).

²¹ Vgl. Meschede 2008, S. 185.

²² Vgl. Jähne 2005, S. 230.

²³ Vgl. RP Photonics Consulting GmbH (Hg.).

²⁴ Vgl. Steinmüller 2008, S. 231.

Raumrichtungen gestreut²⁵. Für eine Lambertsche Oberfläche stehen Veränderungen der reflektierten Lichtintensität eines gegebenen Punktes in direkter Beziehung zum Normalenvektor dieses Punktes. Verläuft die Richtung der Lichtquelle parallel zum Normalenvektor, weist die Oberfläche die größte Helligkeit auf. Besteht zwischen der Richtung der Lichtquelle und dem Normalvektor Orthogonalität, empfängt die Oberfläche kein Licht und ist damit dunkel. Aus diesem Zusammenhang lässt sich die Oberflächenorientierung berechnen. Die Grenze dieser Methode resultiert aus der Tatsache, dass viele Objektoberflächen die Eigenschaft einer Lambertschen Oberfläche lediglich bis zu einem gewissen Grad erfüllen. Insbesondere für metallische oder polierte Materialien kann das Konzept kaum angewendet werden. Weiterhin geht man bei der Beleuchtung von parallel einfallenden Lichtstrahlen, d.h. weit entfernten Lichtquellen aus. Auch dies ist in der Praxis oft, jedoch keineswegs immer gegeben. Weiterführende Informationen finden sich bei [Steinmüller 2008: Bildanalyse].

2.1.2.2 Tiefe aus Textur

Wie zuvor bereits mehrfach erläutert, lassen sich auch anhand des Texturverlaufes einer Oberfläche Aussagen zur Raumtiefe treffen. Die einzelnen Texturelemente einer Fläche werden als Texel (texture elements) bezeichnet²⁶. Ist die Grundform einzelner Texel bekannt, erfolgt für diese aufgrund der perspektivischen Projektion eine Deformierung und eine Größenänderung, aus welcher die Orientierung der Objektoberfläche berechnet werden kann²⁷. In diesem Zusammenhang wird ebenfalls das Konzept des Normalenvektors genutzt. Die Verfahren, die von einer bekannten Texelform ausgehen, werden als strukturelle Verfahren bezeichnet. Sie sind geeignet für Flächen, die regelmäßige und markante Muster aufweisen. Ihnen gegenüber stehen die statistischen Verfahren, welche zur Analyse von natürlich vorkommenden Texturen verwendet werden. Die Analyse der Textur zur Tiefenbestimmung unterliegt einigen Einschränkungen. So werden beispielsweise lediglich Informationen über die Oberflächenorientierung geliefert und keine absoluten Tiefenwerte. Darüber hinaus sind die bekannten Verfahren nur bedingt genau, gleichzeitig aber sehr rechenaufwändig²⁸.

²⁵ Vgl. Nischwitz; Fischer; Haberäcker 2007, S. 166.

²⁶ Vgl. Nischwitz; Fischer; Haberäcker 2007, S. 217.

²⁷ Vgl. Steinmüller 2008, S. 247.

²⁸ Vgl. Ebers 2004, S. 83.

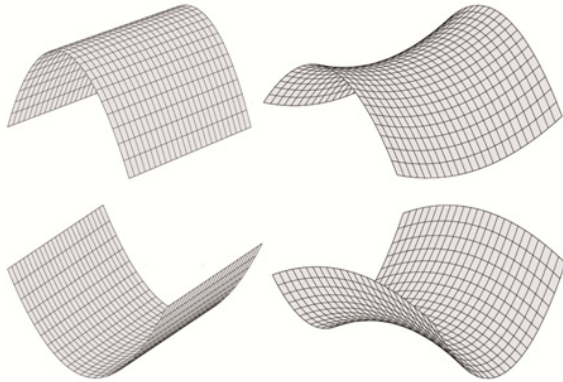


Abbildung 3: Texturverläufe

2.1.2.3 Tiefe aus Schärfe / Unschärfe

Hier werden Aussagen über die Oberflächen von Objekten mit Hilfe der Änderung schärfebeeinflussender Kameraparameter beim Vorhandensein von zwei oder mehr Aufnahmen getroffen. Die Grundlage dafür bildet die im vorherigen Artikel beschriebene Konstruktion der optischen Abbildung. Die Schärfe eines Bildes kann beurteilt werden, indem die Verteilung der Bildenergie über die Frequenzbereiche hinweg betrachtet wird. Hochfrequente Anteile mit starken Koeffizienten weisen dabei auf eine scharfe Abbildung hin²⁹. Bei der Bestimmung von Tiefe aus Schärfe wird der Kamerafokus während der Aufnahme verändert. Dadurch können Beziehungen festgestellt werden zwischen den Szenenobjekten und dem Teil derjenigen Bildpixel, welche scharf abgebildet werden. Wird das Prinzip der Tiefe aus Unschärfe verfolgt, so werden von der Szene nacheinander Aufnahmen mit verschiedenen Schärfereinstellungen erstellt³⁰. Abhängig von den Unterschieden zwischen den Frequenzkoeffizienten der einzelnen Aufnahmen kann so ebenfalls die Tiefenstruktur der sichtbaren Objekte analysiert werden.

2.1.2.4 Tiefe aus Kontur

Silhouetten und Umrisse liefern viele Informationen über die Gestalt und Position eines Objekts und können relativ einfach ermittelt werden. Die dreidimensionale Struktur eines komplexen Objekts kann über eine Kombination mehrerer Konturen erfasst werden. Hierfür wird das Konzept der sog. visuellen Hülle verwendet, es handelt sich dabei um eine aus einfachen Umrissen erzeugte Approximation an die geometrische Form des Objekts³¹. Eine solche Annäherung kann erstellt werden, indem für den Hintergrund und die Objekte einer Szene jeweils unterschiedliche Tiefenebe-

²⁹ Vgl. Ebers 2004, S. 76.

³⁰ Vgl. Niederoest; Niederoest; Šcucka 2003.

³¹ Vgl. Ebers 2004, S. 68.

nen definiert werden. Bildbereiche, die von Änderungen betroffen sind, z.B. Bewegungen von Objekten, können über die Berechnung der Unterschiede zwischen zwei aufeinanderfolgenden Bildern ermittelt werden. Dabei werden all diejenigen Pixel dem Vordergrund zugeordnet, welche erhebliche Intensitätsänderungen aufweisen.

2.1.2.5 Tiefe aus Bewegung

Dieses Verfahren kann für zwei unterschiedliche Fälle eingesetzt werden. Im ersten Fall sind eine statische Kamera und sich bewegende Szenenobjekte gegeben. Dabei werden auffällige Merkmalspunkte bei den Objekten erkannt und deren Bewegung über eine Abfolge von Bildern hinweg verfolgt. Damit läuft das Problem auf den Vergleich der Unterschiede zweier Bilder hinaus. Dies wird als Korrespondenzproblem bezeichnet, welches im Zusammenhang mit der Gewinnung von Tiefe über Stereotriangulation im nächsten Abschnitt behandelt wird. Im zweiten Fall bewegt sich die Kamera, um dadurch – ähnlich wie bei der Stereoskopie – eine Szene aus unterschiedlichen Perspektiven zu erfassen. Eine dreidimensionale Bewegung kann so mit Hilfe von Verschiebungsvektoren annähernd beschrieben werden³².

2.1.3 Tiefe aus Stereotriangulation

Dieses Verfahren zählt im Grunde zu den passiven Verfahren, es wird allerdings eine zusätzliche Kamera eingesetzt. Hierbei wird die Tatsache, dass ein Objekt aus zwei horizontal unterschiedlichen Positionen aufgenommen wird, ausgenutzt. Dies hat zur Folge, dass das Objekt, bzw. seine Punkte auch auf unterschiedlichen Positionen in der Bildebene abgebildet werden. Die Differenz zwischen den Positionen auf der Bildebene wird, wie bereits mehrfach erwähnt, als Parallaxe oder Disparität bezeichnet. Ist die Position der Bildpunkte bekannt, so kann mit diesen Werten sowie der Brennweite und der Stereobasis die Disparität berechnet werden, welche wiederum umgekehrt proportional zur Entfernung des Objektpunktes ist. Das eigentliche Problem bei der Stereotriangulation ist, in beiden Stereohalbbildern die beiden Punkte zu finden, die die Abbildung desselben Objektpunktes darstellen. Dieses spezifische Problem wird als Stereoanalyse bezeichnet und basiert auf der Korrespondenzanalyse, einer bekannten Aufgabe aus der Bildverarbeitung und insbesondere der Bildcodierung³³.

2.2 Gewinnung von Tiefeninformation aus Stereobildern

Jedes Prinzip der Tiefengewinnung bietet spezifische Vor- und Nachteile. Die Nutzung eines bestimmten Verfahrens hängt oft von Einsatzgebiet, Anforderungen und verfügbaren Ressourcen ab. Im Hinblick auf die bisher in dieser Arbeit gelieferten Informationen bietet sich bei der S3D-Produktion die Gewinnung von Tiefeninformation mittels Stereotriangulation an, da eine der Bedingungen dafür der Einsatz eines binokularen Kameraaufbaus ist. Somit entfällt die zusätzliche Implementierung von

³² Vgl. Steinmüller 2008, S. 253.

³³ Vgl. Waldowski (Hg.) 2010, S. 54ff.

Hard- und Softwarekomponenten, die u.U. für die Nutzung eines anderen Verfahrens notwendig ist.

Tiefeninformationen lassen sich aus Stereobildpaaren über die sog. Stereoanalyse extrahieren. Die Stereoanalyse ist ein Untergebiet der Korrespondenzanalyse. Das Ziel der Korrespondenzanalyse in der Bildverarbeitung ist allgemein der Vergleich von Bildern zur Ermittlung von Gemeinsamkeiten.

2.2.1 Fehlerquellen bei der Stereoanalyse

In der Literatur werden einige generelle Sachverhalte beschrieben, die bei Stereoaufnahmen auftreten können und sich problematisch auf die Korrespondenzanalyse auswirken³⁴. Es kann beispielsweise Fälle geben, bei denen für bestimmte Bildinhalte keine Korrespondenzen ermittelt werden können, da die Inhalte lediglich in einem der beiden Stereohalb Bilder vorkommen. Weitere Probleme ergeben sich durch spezifische Bildmuster. Im Einzelnen müssen bei der Stereoanalyse eventuell für folgende Probleme Lösungen gefunden werden:

- Verdeckung: In einem der beiden Halbbilder wird ein Teil des Objekts A durch Objekt B verdeckt. Aufgrund des perspektivischen Unterschiedes ist im anderen Halbbild dieser Teil von A erkennbar, möglicherweise wird jedoch gleichzeitig ein anderer Bereich verdeckt. Damit können für beide Bereiche keine Korrespondenzen hergestellt werden, da sie lediglich in je einem der beiden Bilder vorkommen³⁵.
- Unterschiedliche Sichtbereiche der Kameras: Jeweils eine Kamera kann aufgrund der unterschiedlichen Perspektive nicht den exakt gleichen Szenenausschnitt erfassen wie die andere Kamera. Konkret kann die rechte Kamera nicht den linken Randbereich des linken Bildes erfassen und die linke Kamera nicht den rechten Randbereich des rechten Bildes.
- Periodische Motivstrukturen: Hier tritt ein markantes Muster mehrmals und in vielen Fällen auch in benachbarten Bildregionen auf. Die Muster an sich können zwar relativ einfach erkannt werden, jedoch verläuft die Positionszuordnung der Korrespondenzen fehlerhaft. So werden vermeintlich korrekte Verbindungen zwischen Bildbereichen hergestellt, welche de-facto nicht zusammengehören³⁶.
- Homogene Motivbereiche: Hier handelt es sich um Motive, die nach der Aufnahme als einheitliche Fläche abgebildet werden. So können keine für die Korrespondenzanalyse notwendigen Bildmerkmale wie beispielsweise Helligkeits-

34 Vgl. Schreer 2005, S. 127.

35 Vgl. Klaus; Sormann; Kärner 2006.

36 Vgl. Schreer 2005, S. 127.

oder Farbunterschiede erkannt und verglichen werden. Dieses Problem tritt bei realen Kameraaufnahmen insbesondere auch im Vordergrund oder Hintergrund auf, wenn der Schärfentiefebereich der Optik über- bzw. unterschritten wird.

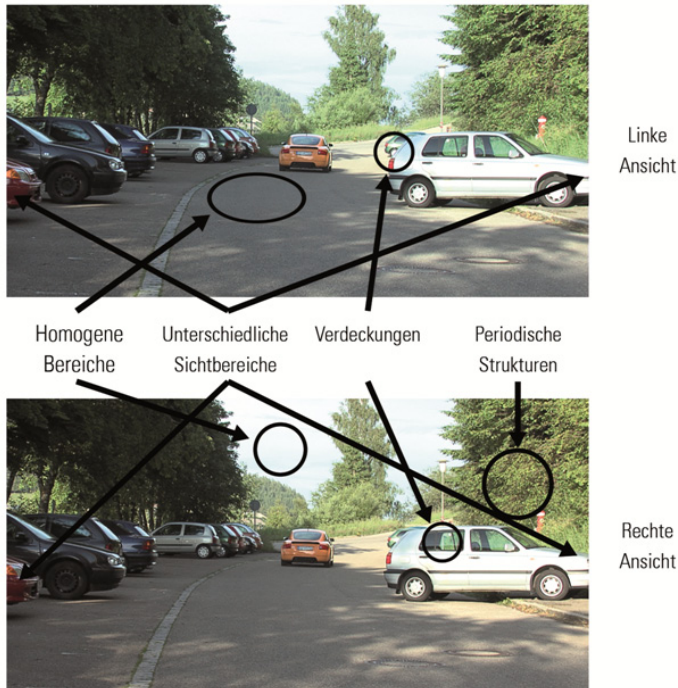


Abbildung 4: Problembereiche der Korrespondenzanalyse

2.2.2 Genauigkeit der Stereoanalyse

Die Genauigkeit einer Tiefenkarte kann über die gefundenen Korrespondenzen definiert werden. Können für jeden Objektpunkt die beiden zusammengehörenden Abbildungen in linker und rechter Ansicht eindeutig zugeordnet werden, beträgt die Genauigkeit 100 Prozent. Dieser Fall ist jedoch i. A. so gut wie niemals gegeben.

2.2.2.1 Ortsauflösung

Zunächst muss - selbst bei optimalen Fokalverhältnissen - eine generelle Einschränkung hinsichtlich der Genauigkeit einer Tiefenkarte erläutert werden. Diese Einschränkung besteht in der begrenzten Fähigkeit zur Ortsauflösung einer Oberfläche

abhängig von der Bildrasterung und der Entfernung zwischen Kamera und Objekt³⁷. Ein Objektpunkt, dessen Abbildung die räumlichen Dimensionen eines Pixels unterschreitet, kann nicht mehr als Punkt erfasst werden. Details können daher ab einem bestimmten Grad nicht aufgenommen und somit auch nicht rekonstruiert werden. Für die maximal erreichbare Ortsauflösung ergibt sich aus den Größen Pixelhöhe, bzw. –breite p , Brennweite f und Objektweite a folgende Grenzbetrachtung³⁸:

$$\frac{p}{f} = \frac{\text{erreichbare Auflösung}}{a}$$

Formel 1

2.2.2.2 Genauigkeitsrelevante Parameter

Der Grad der Genauigkeit bei der Erstellung einer Tiefenkarte aus einem digitalen Videobild hängt von mehreren verschiedenen Faktoren ab, welche sich wiederum teilweise gegenseitig beeinflussen. Hierzu zählen die Eigenschaften des Videosignales selbst, die verwendeten Soft- und Hardwarekomponenten und die Bildinhalte, repräsentiert durch die Helligkeits- und Farbwerte der Pixel³⁹. Ein weiteres entscheidendes Element dieser Konstellation stellt der Faktor der Verarbeitungszeit dar.

- Pixelraster: Je mehr Pixel ein Bild umfasst, desto mehr Punkte müssen zwischen linker und rechter Ansicht verglichen werden.
- Aufnahmezeit: Je mehr Einzelbilder in einer Zeitperiode aufgenommen werden, desto weniger Zeit steht für die Berechnung der Tiefenkarte eines einzelnen Bildes zur Verfügung.
- Bildinhalte: Wie bereits erläutert, kann es algorithmisch sehr schwer werden, für bestimmte Pixelwerte in einer Ansicht die zugehörige Korrespondenz in der anderen Ansicht zu ermitteln. Dies hängt mit den Methoden zusammen, über die die Pixelwerte verglichen werden. Da die Vergleichsverfahren i.d.R. auf die Helligkeits- und/oder Farbinformationen zurückgreifen, verursachen Bildbereiche mit einheitlichen Helligkeitsstufen und ähnlichen Farbtönen Probleme, da benachbarte Pixel gleiche oder sehr ähnliche Werte aufweisen. Auch Bildfehler wie z.B. Rauschen oder Kompressionsartefakte, müssen zunächst als solche identifiziert und dann entweder korrigiert oder für den Vergleich ignoriert werden.
- Hard- und Software: Abhängig davon, wie viele mathematischen Operationen ein Computersystem innerhalb einer Zeitperiode durchführen kann, kann eine

³⁷ Vgl. Steinmüller 2008, S. 203.

³⁸ Vgl. Steinmüller 2008, S. 203.

³⁹ Vgl. Hile; Zheng 2004.

Tiefenkarte nur mit einem bestimmten Genauigkeitsgrad innerhalb einer bestimmten Zeit erstellt werden. Bei den Algorithmen, über die die Operationen durchgeführt werden, herrscht eine große Vielfalt. Viele der Programme sind für bestimmte Anforderungen optimiert und variieren daher in der Performance bei gleichen Hardware-Ressourcen.

Der optimale Fall stellt die maximale Genauigkeit, maximale Pixelrastergröße, maximale Framerate und maximale Helligkeits- und Farbwiedergabe bei minimaler Verarbeitungszeit und minimal notwendigen Hard- und Softwareressourcen dar. Dabei soll eine Unabhängigkeit vom Bildinhalt gegeben sein. In den meisten Situationen existieren bestimmte Mindestanforderungen für einen oder mehrere Parameter. Mit diesen Anforderungen als Ausgangslage werden die verbleibenden Parameter solange manipuliert, bis die Anforderungen erfüllt sind oder das bestmögliche Ergebnis erzielt wurde.

Beispiel: Für die Tiefenkarte eines gegebenen Motivs muss ein gewisser Grad an Genauigkeit gewährleistet sein und nun soll die notwendige Verarbeitungszeit reduziert werden. Als mögliche Optionen können dann

- die räumliche Auflösung verringert,
- weniger Einzelbilder pro Sekunde aufgenommen,
- die Quantisierungsstufen verringert,
- leistungsfähigere Hardware-Komponenten organisiert,
- Verarbeitungsalgorithmen optimiert oder
- mehrere der genannten Optionen kombiniert werden.

Je nach Anwendung besitzen unterschiedliche Parameter unterschiedliche Prioritäten. Aus diesem Grund handelt es sich bei der praktischen Umsetzung eines solchen Systems meist um einen Kompromiss zwischen den verfügbaren Ressourcen und dem angestrebten Ziel. Für Aufnahmen von Film- und Videomaterial bestehen feste Anforderungen für die Bereiche Pixelrastergröße, Framerate, Helligkeits- und Farbtreue, bei gleichzeitig gewünschter Unabhängigkeit von den Inhalten des Bildes. Als manipulierbare Parameter für die Genauigkeit der Tiefenkarte verbleiben daher theoretisch die Hard- und Software sowie der zur Verfügung stehende Berechnungszeitraum⁴⁰. Soll die Nutzung der Tiefenkarte in Echtzeit erfolgen, so reduzieren sich die variablen Parameter theoretisch auf Hard- und Software sowie algorithmische Optimierungen zur Laufzeit.

2.2.3 Ähnlichkeitsbedingungen bei der Stereoanalyse

Für ein Stereokamerasystem können verschiedene Einschränkungen definiert werden,

40 Anm.: Bei eher statischen Szenen kann für die Tiefenberechnung der Frame-Input natürlich doch reduziert werden.

um die Zuverlässigkeit der Korrespondenzanalyse zu verbessern, bzw. die Berechnung zu vereinfachen. Diese Einschränkungen können z.T. aus der Bildgeometrie abgeleitet werden oder sie resultieren aus Vorwissen über die Objekte im Bild.

- Epipolarbedingung (epipolar constraint): Diese Bedingung wurde bereits im vorangegangenen Artikel umfassend erläutert. Zwei korrespondierende Punkte müssen auf korrespondierenden Epipolarlinien liegen. Diese verlaufen nach der Rektifikation horizontal und entsprechen den Bildzeilen.
- Eindeutigkeitsbedingung (uniqueness constraint): Ein Pixel in einer Ansicht kann nur mit einem Pixel der anderen Ansicht korrespondieren, wenn sie nicht von einer Verdeckung betroffen sind⁴¹.
- Reihenfolgebedingung (ordering constraint): Bei der Reihenfolge von Mustern entlang der Epipolarlinie einer Ansicht ist normalerweise davon auszugehen, dass diese Reihenfolge für die Epipolarlinie der zweiten Ansicht erhalten bleibt. Bei starken perspektivischen Unterschieden bezüglich der Position der Kameras zueinander und der Tiefenstruktur der Szene kann diese Bedingung allerdings verletzt werden (siehe Abb. 5)⁴².
- Verträglichkeit von Intensitätswerten (photometric compatibility constraint): Bezüglich der Pixelwerte, die für die Analyse herangezogen werden, ist es wahrscheinlich, dass zwei korrespondierende Pixel ähnliche Helligkeits- und Farbwerte aufweisen⁴³.
- Geometrische Ähnlichkeit der Merkmale (geometric similarity constraint): Werden bestimmte Merkmale ermittelt, so können deren geometrische Eigenschaften überprüft werden. Beispielsweise kann von zwei korrespondierenden Objektsegmenten, die die Form von Linien haben, angenommen werden, dass sie sich in Länge und Orientierung nicht sehr stark voneinander unterscheiden⁴⁴.
- Glattheitsbedingung (disparity smoothness constraint): Bei dieser Bedingung wird davon ausgegangen, dass sich die Disparitätswerte benachbarter Pixel lediglich innerhalb eines gewissen Rahmens ändern⁴⁵. Probleme treten allerdings auf, wenn sich die Tiefe abrupt ändert, z.B. an Kanten.
- Kontinuität entlang der Kanten (figural disparity constraint): Der korrespondierende Punkt für einen Pixel, welcher in der linken Ansicht auf einem Kantenelement liegt, findet sich in der rechten Ansicht mit hoher Wahrscheinlichkeit ebenfalls entlang eines Kantenelements und umgekehrt⁴⁶.

41 Vgl. Zhan; Lee; Beard 2004.

42 Vgl. Steinmüller 2008, S. 230.

43 Vgl. Zhan; Lee; Beard 2004.

44 Vgl. Scharfenberg 2009.

45 Vgl. Faugeras et al. 1993, S. 3.

46 Vgl. Zhan; Lee; Beard 2004.

- Verträglichkeit der Merkmale (feature compatibility constraint): Zwischen zwei Bildpunkten kann nur dann eine Korrespondenz bestehen, wenn sie den gleichen physikalischen Objektpunkt als Ursprung haben⁴⁷.
- Disparitätsgradientenlimit (disparity limit constraint): Liegt hinreichend Kenntnis über die Tiefe der Szene vor, beispielweise Nah- und Fernpunkt, so kann aufgrund der reziproken Beziehung zwischen Raumtiefe und Disparität ein maximal möglicher Disparitätswert definiert werden⁴⁸. Überschreitet ein ermittelter Disparitätswert dieses Maximum deutlich, so handelt es sich dabei um eine falsche Zuordnung.

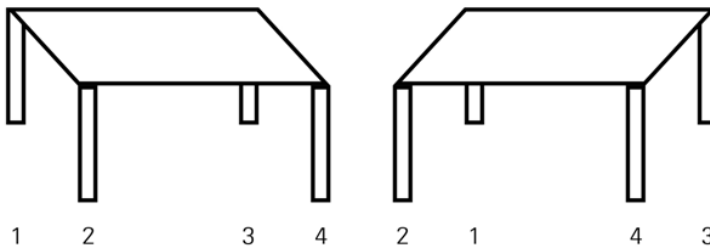


Abbildung 5: Reihenfolgebedingung

2.2.4 Pixelbasierte Verfahren

Bei pixelbasierten Verfahren werden zur Korrespondenzanalyse die Intensitäts- und ggf. Farbwerte eines Pixels herangezogen. Allerdings werden für einen bestimmten Pixel nicht nur dessen Werte genutzt, sondern auch die Werte der umliegenden Bildpunkte. Es werden also Pixelblöcke untersucht, diese Vorgehensweise wird daher Block-Matching genannt. Für einen Pixel A an der Position (u_i, v_i) wird in einer Stereoansicht ein Referenzblock der Größe (m, n) gebildet. Dieser Block wird mit gleich-großen Musterblöcken in der zweiten Stereoansicht hinsichtlich ihrer Ähnlichkeit verglichen. Der Musterblock, der die größte Ähnlichkeit aufweist, wird dem Referenzblock zugeordnet, woraus sich eine relative Positionsverschiebung zwischen den Blöcken errechnen lässt. Bei dieser Verschiebung handelt es sich um die Disparität δ . Im Allgemeinen besitzt die Disparität eine vertikale und horizontale Dimension $\delta(u, v)$. Handelt es sich allerdings um rektifizierte Ansichten, so verbleibt lediglich die horizontale Komponente. Zum Vergleich der Pixelblöcke können mathematische Operationen verwendet werden, in der Literatur werden diese unterteilt in parametrische und nicht-parametrische Ähnlichkeitsmaße⁴⁹.

47 Vgl. Steinmüller 2008, S. 230.

48 Vgl. Steinmüller 2008, S. 228.

49 Vgl. Schreer 2005, S. 134 & 138.

2.2.4.1 Parametrische Ähnlichkeitsmaße

Mittlerer absoluter Fehler (mean absolute difference, kurz: MAD): Hierbei wird die absolute Differenz zwischen zwei Pixelblöcken berechnet, indem die Pixelwerte der Blöcke voneinander subtrahiert und betragsweise aufsummiert werden. Der Ergebniswert ist immer ≥ 0 und die Ähnlichkeit zwischen den Blöcken ist am größten bei der kleinsten absoluten Differenz⁵⁰.

$$\delta_{opt}(u, v) = \underset{\delta(u, v)}{\operatorname{argmin}} \frac{1}{|\Lambda|} \sum_m \sum_n |f_1(u + m, v + n) - f_2(u + \delta(u, v) + m, v + n)|$$

Formel 2

Für den mittleren absoluten Fehler kann zusätzlich eine sog. Mittelwertbefreiung durchgeführt werden. Der Hintergrund dabei ist, dass durch unterschiedliche Blendendurchmesser bei den Kameras Unterschiede bei der mittleren Helligkeit zwischen beiden Ansichten auftreten können. Wird dieses Problem auf Aufnahmen für Film und Video übertragen, werden bei sorgfältiger Arbeitsweise kaum Unterschiede zwischen den Blendeneinstellungen auftreten. Kommt allerdings ein Mirror-Rig zum Einsatz, können Helligkeitsunterschiede in den Ansichten durch den Strahlenteiler verursacht werden, daher kann die Mittelwertbefreiung zur Verbesserung der Korrespondenzanalyse durchaus sinnvoll sein. Die Komponenten von Formel 121 müssen hierzu erweitert werden⁵¹:

$$(f_1(u + m, v + n) - \bar{f}_1), (f_2(u + \delta(u, v) + m, v + n) - \bar{f}_2)$$

$$\bar{f}_1 = \frac{1}{|\Lambda|} \sum_m \sum_n^{\text{mit}} f_1(u + m, v + n)$$

und

$$\bar{f}_2 = \frac{1}{|\Lambda|} \sum_m \sum_n f_2(u + \delta(u, v) + m, v + n)$$

Formel 3

• Mittlerer quadratischer Fehler (mean square error, kurz: MSE): Dieses Vergleichsverfahren minimiert die Energie der Bilddifferenz. Der quadratische Fehler liefert dann den kleinsten Wert, wenn die Ähnlichkeit zwischen den zu vergleichenden Pixelblöcken am größten ist⁵².

⁵⁰ Vgl. Waldowski (Hg.) 2010, S. 54f.

⁵¹ Vgl. Schreer 2005, S. 136.

⁵² Vgl. Schreer 2005, S. 135.

$$\delta_{opt}(u, v) = \underset{\delta(u, v)}{\operatorname{argmin}} \frac{1}{|\Lambda|} \sum_m \sum_n [f_1(u + m, v + n) - f_2(u + \delta(u, v) + m, v + n)]^2$$

Formel 4

• Normierte Kreuzkorrelation (normalized cross-correlation, kurz: NCC): Hier werden die Werte der Pixelblöcke in Form von zwei Vektoren **a** und **b** dargestellt. Die Kreuzkorrelation entspricht dann dem Skalarprodukt aus **a** mit **b**. Damit die Vektoren unabhängig von ihrer Länge verglichen werden können, müssen sie auf die Länge 1 normiert werden. Je geringer dann der Betrag der Vektordifferenz ist, desto ähnlicher sind sich **a** und **b**⁵³.

$$\delta_{opt}(u, v) = \underset{\delta(u, v)}{\operatorname{argmax}} \frac{\sum_m \sum_n f_1(u + m, v + n) - f_2(u + \delta(u, v) + m, v + n)}{\sqrt{\sum_m \sum_n (f_1(u + m, v + n))^2 * \sum_m \sum_n (f_2(u + \delta(u, v) + m, v + n))^2}}$$

Formel 5

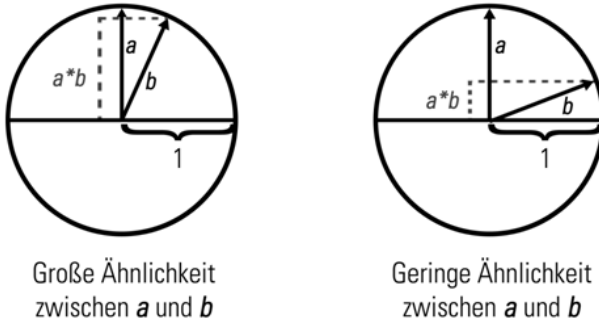


Abbildung 6: Normierte Kreuzkorrelation

2.2.4.2 Nicht-parametrische Ähnlichkeitsmaße

Über nicht-parametrische Ähnlichkeitsmaße lassen sich nicht die Intensitäten einzelner Pixel vergleichen, sondern deren relative Anordnung zueinander. Eine entsprechende Transformation wird i. A. durchgeführt, bevor der Vergleich über eines der im vorherigen Abschnitt erläuterten Verfahren (MAD, MSE, NCC) erfolgt. Zu diesen Transformationen zählen u.a. die Rank-Transformation und die Census-Transformation⁵⁴. Das Ziel dieser Verfahren besteht darin, die parametrischen Ähnlichkeitsmaße weniger anfällig für Probleme durch Disparitätssprünge zu machen, wenn also die Werte innerhalb der Pixelblöcke große Unterschiede aufweisen.

⁵³ Vgl. Psarakis; Evangelidis 2005.

⁵⁴ Vgl. Zabih; Woodfill 1994.

- **Rank-Transformation:** Mit dieser Transformation lässt sich für jeden Pixel A innerhalb eines bestimmten Bereichs mit N Pixeln die Anzahl der Bildpunkte berechnen, deren Intensität kleiner ist als die von A. Auf diese Weise ergibt sich für jeden Pixel ein neuer Wert, welcher innerhalb eines Bereichs von 0 bis N-1 liegt⁵⁵. Für beide Bilder des Stereopaars wird eine Rank-Transformation durchgeführt und anschließend werden die Pixel über ein parametrisches Ähnlichkeitsmaß verglichen.
- **Census-Transformation:** Bei der Census-Transformation wird die Intensität von Pixeln über ein Bit-Wort beschrieben⁵⁶. Dies geschieht ähnlich wie bei der Rank-Transformation auf eine Weise, bei der N Pixelwerte innerhalb eines Bereichs relativ zur Intensität eines bestimmten Pixels A betrachtet werden. Die Länge des Bitwortes entspricht dabei der Anzahl der zu vergleichenden Pixel des Messbereichs. Die Ähnlichkeit zwischen zwei ausgewählten Bit-Worten kann dann über das Prinzip der sog. Hamming-Distanz ermittelt werden⁵⁷. Aus der Summe der Hamming-Distanzen wird anschließend die Ähnlichkeit zwischen zwei Pixelblöcken ebenfalls über ein parametrisches Ähnlichkeitsmaß berechnet.

2.2.4.3 Methoden zur Bestimmung der optimalen Blockgröße

Bei der Anwendung von Verfahren, die auf Block-Matching beruhen, tritt schnell das Problem der passenden Blockgröße auf. Um Disparitäten zuverlässig ermitteln zu können, sind große Intensitätsvariationen bei den Pixeln innerhalb des Blockes hilfreich. Um eine entsprechende Vielfalt erreichen zu können, müssen ausreichend große Blöcke verwendet werden. Dann jedoch besteht die Gefahr, dass die Blöcke Bereiche beinhalten, in denen es zu Tiefensprüngen und somit auch zu Disparitätssprüngen kommt. In diesem Fall kann möglicherweise nicht für alle Pixel eine Korrespondenz ermittelt werden. Bei einer kleinen Blockgröße wird das Problem der Tiefensprünge minimiert, entsprechend reduziert sich dann jedoch auch die Intensitätsvielfalt innerhalb des Blocks. Somit besteht ein Zusammenhang zwischen der Genauigkeit, mit welcher die Disparität bestimmt wird und der Zuverlässigkeit bei der Zuordnung von Korrespondenzen. Im Folgenden werden Methoden beschreiben, die unterschiedliche Lösungsansätze für das Problem der optimalen Blockgröße verfolgen.

- **Gaußpyramide:** Eine Gaußpyramide entsteht, wenn aus einem Startbild über Unterabtastung mehrere Bilder mit geringerer Auflösung gewonnen werden⁵⁸. Für das Block-Matching geht man für die linke und rechte Stereoansicht von den Bildern mit der kleinsten Auflösung aus und stellt die Korrespondenzen innerhalb eines festen Suchbereiches her. Das Ergebnis dient als Startwert für

⁵⁵ Vgl. Zabih; Woodfill 1994.

⁵⁶ Vgl. Zabih; Woodfill 1994.

⁵⁷ Vgl. Hamming 1950.

⁵⁸ Vgl. Jähne 2005, S. 147.

die Bilder mit der nächsthöheren Auflösung, wobei sich der Suchbereich im Verhältnis zur Bildgröße verkleinert⁵⁹. Diese Iteration wird für die verschiedenen Auflösungsstufen durchlaufen und führen letztlich im Originalbild zu einem sehr kleinen Pixelblock, für den schnell die korrekten Korrespondenzen ermittelt werden können.

- **Coarse-to-Fine:** Hier wird bei voller Bildauflösung eine Korrespondenzsuche durchgeführt. Der Suchbereich und die Blockgröße werden dabei in mehreren Durchläufen reduziert⁶⁰.

- **Zweistufiges Blockmatching:** Diese Methode beinhaltet eine lokale und eine globale Analyse. Im globalen Schritt werden einige Merkmalspunkte bestimmt, welche über den gesamten Bildbereich verteilt sind, um die Tiefenstruktur der Szene möglichst vollständig zu repräsentieren. Für diese Punkte wird eine Korrespondenzanalyse durchgeführt, dabei wird ein großer Suchbereich und ein großes Suchfenster genutzt. Das Ergebnis wird anschließend auf der lokalen Stufe als Ausgangspunkt zur Bestimmung der Korrespondenzen für die einzelnen Pixel verwendet. Da die Disparitäten aus der globalen Analyse vorliegen, können der Suchbereich und das Suchfenster entsprechend verkleinert werden.

- **Adaptive Fenster:** Hier werden die Messfenster an den Bildinhalt angepasst. Das Konzept der adaptiven Fenster sieht vor, dass für jeden Bildpunkt hinsichtlich seiner Form und Größe ein geeigneter Pixelbereich ermittelt wird⁶¹. Dabei sind die Anforderungen an den Bereich erneut eine große Vielfalt an Intensitätswerten und kein Vorkommen von Tiefsprüngen. Die Variation der Intensitätswerte kann über die bereits genannten Verfahren berechnet werden. Das Auftreten von Tiefsprüngen wird verhindert, indem die Form des Messfensters anhand eines statistischen Modells für die Disparität iterativ angepasst wird. Bei einer alternativen Vorgehensweise werden für jeden Pixel neun Fenster mit unterschiedlichen Aufpunkten analysiert⁶². Es wird dabei angenommen, dass das Fenster, in welchem keine Disparitätssprünge auftauchen, zu einem optimalen Ergebnis bei der Berechnung der Ähnlichkeit führt.

2.2.4 4 Hybrid-Rekursives Matching

Dieses Verfahren verfolgt nicht den Ansatz einer Suche in einem Bereich entlang der Epipolarlinie. Das hybrid-rekursive Matching beinhaltet eine block-rekursive Disparitätsschätzung und eine pixel-rekursive Schätzung des optischen Flusses⁶³. Es kann daher lediglich bei Sequenzen aufeinanderfolgender Bilder genutzt werden. Für den ersten Schritt der Block-Rekursion werden drei Blöcke mit einem Referenzblock ver-

59 Vgl. Schreer 2005, S. 140.

60 Vgl. Nillius; Eklundh 2002.

61 Vgl. Kanade; Okutomi 1991.

62 Vgl. Bobick; Intille 1999.

63 Vgl. Schreer 2005, S. 144.

glichen. Es handelt sich dabei um zwei örtlich benachbarte Blöcke und den Block, welcher im vorhergehenden Bild an der Position des Referenzblocks liegt. Für diese Elemente wird über die Differenz verschobener Blöcke (displaced block difference, kurz: DBD) die Ähnlichkeit mit dem Verfahren des mittleren absoluten Fehlers berechnet:

$$\text{DBD}(\mathbf{d}) = \sum_{x=0}^M \sum_{y=0}^N |f_1(u, v) - f_2(u + d_u, v + d_v)|$$

Formel 6

Es ergibt sich ein Vektor, welcher die Startposition für die Pixel-Rekursion liefert. Mit Hilfe von örtlichen Gradienten und dem Gradienten zwischen linkem und rechtem Stereobild wird ein Update-Vektor \mathbf{d} berechnet. Der Gradient zwischen den Stereobildern lässt sich nach Schreer über die Differenz verschobener Pixel (displaced pixel difference, kurz: DPD) ermitteln⁶⁴:

$$\mathbf{d}(u, v) = \mathbf{d}_i - \text{DPD}(\mathbf{d}_i, u, v) * \frac{\text{grad } f_1(u, v)}{\|\text{grad } f_1(u, v)\|^2} \text{ mit}$$

$$\text{DPD}(\mathbf{d}_i, u, v) = |f_1(u, v) - f_2(u + d_u, v + d_v)|$$

Formel 7

Der Vektor \mathbf{d} dient als Startwert für eine neue Berechnung an der nächsten Pixelposition innerhalb des Blocks. Wurde der gesamte Block durchlaufen, wird der Update-Vektor ausgewählt, welcher die kleinste DPD liefert. Danach wird dieser Vektor mit dem Vektor aus der Block-Rekursion verglichen und der kleinere der beiden als endgültiger Vektor übernommen. Auf diese Weise ermöglicht die Pixel-Rekursion die schnelle Anpassung an eine neue Tiefenstruktur, während die Block-Rekursion eine örtliche und zeitliche Stabilisierung der Tiefenkarte gewährleistet.

Um korrespondierende Disparitätsvektoren zu überprüfen, kann ein sog. Konsistenztest durchgeführt werden. Wenn die Disparitätskarten für Bildpunkte von Ansicht 1 in Ansicht 2 ($1 > 2$) und umgekehrt ($2 > 1$) vorliegen, so können die entsprechenden Vektoren verglichen werden. Hierzu wird ein Schwellenwert definiert, den die Abweichung zwischen zwei Disparitätsvektoren nicht überschreiten darf.

2.2.4.5 Dynamische Programmierung

Bei dieser Methode zur Durchführung der Korrespondenzanalyse wird das Zuordnungsproblem in mehrere kleine Teilbereiche aufgeteilt. Dabei wird zunächst für alle

⁶⁴ Vgl. Schreer 2005, S. 146.

paarweisen Korrespondenzen zwischen den Pixeln auf korrespondierenden Zeilen in den beiden Stereoansichten das Ähnlichkeitsmaß berechnet⁶⁵. Daraufhin werden die Ähnlichkeitsfunktion schrittweise abgearbeitet und die tatsächlichen Korrespondenzen ausgewählt. Hierzu wird innerhalb der zweidimensionalen Matrix der Ähnlichkeitsfunktion der Weg mit der größten Ähnlichkeit gesucht⁶⁶. Auf diese Weise werden sowohl die Ähnlichkeiten ($1 > 2$) als auch ($2 > 1$) analysiert, wodurch mehrdeutige Zuweisungen ausgeschlossen werden.

2.2.5 Merkmalsbasierte Verfahren

Bei merkmalsbasierten Verfahren werden ausgewählte Bildmerkmale bezüglich ihrer Korrespondenz untersucht. Dies führt dazu, dass weniger Elemente als bei pixelbasierten Verfahren verglichen werden müssen, es jedoch gleichzeitig notwendig wird, die Merkmale in einer Vorverarbeitungsstufe zu ermitteln. Meist werden die Merkmale unterteilt in Punktmerkmale und Liniensegmente⁶⁷.

2.2.5.1 Punktmerkmale

Um die Korrespondenzanalyse effizient durchführen zu können, ist es ratsam, die Menge der Merkmalspunkte angemessen einzuschränken. Die Punkte sollten möglichst über den gesamten Bildbereich verteilt sein, um die Tiefenstruktur der Szene gut erfassen zu können. Besonders markante Punkte stellen Eckpunkte von Kanten dar, da sie eine Intensitätsänderung in zwei Richtungen aufweisen.

- **Bestimmung:** Zur Bestimmung von Eckpunkten existieren verschiedene Verfahren, die als Interest-Operatoren bezeichnet werden. Eines dieser Verfahren ist der sog. Moravec-Operator⁶⁸. Hier wird die mittlere Änderung der Intensität um einen Bildpunkt analysiert. Der Operator liefert dabei den größten Wert für Positionen, an denen die Intensitätsänderung besonders groß ist (d.h. an Ecken). Eine verbesserte Version des Moravec-Operators stellt der Plessey-Punkt-Detektor (auch Harris-Eckendetektor) dar. Hier wird über die Helligkeitsgradienten in horizontaler und vertikaler Richtung entschieden, ob ein Punkt für die Korrespondenzanalyse geeignet ist. Hierzu wird eine Autokorrelations-Matrix A verwendet⁶⁹. Je nach Umgebung des gewählten Pixels besitzt die Matrix einen unterschiedlichen Rang, dabei gilt:

- $Rg(A) = 0$: die Umgebung weist eine gleichmäßige Intensität auf, die Fläche ist homogen und unstrukturiert
- $Rg(A) = 1$: es wurde eine Kante erkannt
- $Rg(A) = 2$: es wurde ein Eckpunkt erkannt

65 Vgl. Leung; Appleton; Sun 2004.

66 Vgl. Schreer 2005, S. 152.

67 Vgl. Schreer 2005, S. 152.

68 Vgl. Moravec 1981.

69 Vgl. Harris; Stephens 1988.

- Korrespondenzanalyse: Wenn in beiden Ansichten Merkmalspunkte ermittelt wurden, kann auf diese eine Korrespondenzanalyse angewandt werden. Hierzu kann die Epipolarbedingung genutzt werden. Es werden zunächst die Korrespondenzen für Punkte der linken Ansicht hergestellt, für die lediglich ein möglicher Gegenpart auf der Epipolarlinie der rechten Ansicht existiert. Zur Erhöhung der Zuverlässigkeit können zusätzlich die Pixelstrukturen um die ermittelten Punkte über einen Suchbereich verglichen werden. Danach verbleiben in der linken Ansicht die Punkte, für die mehrere mögliche Zuordnungen getroffen werden können. Um für diese Kandidaten entsprechende Eindeutigkeiten herstellen zu können, werden zusätzliche Bedingungen der Stereoanalyse herangezogen, z.B. die Reihenfolgebedingung oder die Glattheitsbedingung. Je mehr Merkmalspunkte erkannt und zugeordnet werden, desto genauer wird die Tiefenstruktur erfasst, gleichzeitig erhöht sich allerdings auch der Rechenaufwand.

2.2.5.2 Liniensegmente

Die Identifikation von Liniensegmenten in einem oder mehreren Bildern bringt einen großen Berechnungsaufwand mit sich. Dem gegenüber steht eine vereinfachte Korrespondenzanalyse, da für weniger Elemente Zuordnungen hergestellt werden müssen. Bei der Bestimmung von Liniensegmenten werden die Kanten von Szenenobjekten gesucht, von denen angenommen wird, dass sie einen auffälligen abrupten Übergang von einer hellen zu einer dunkleren Bildregion aufweisen. Bei diesem Verfahren muss jedoch beachtet werden, dass organische, natürlich vorkommende Objekte i.d.R. weniger gerade Kanten aufweisen als beispielsweise künstliche Objekte wie Gebäude oder Möbelstücke. Demzufolge kann der Erfolg einer linienbasierten Korrespondenzanalyse von der Struktur der Szene abhängen, die aufgenommen wird. Die Liniensegmente werden i. A. ermittelt, indem zunächst Kantenpunkte identifiziert und diese dann Linien zugeordnet werden.

- Bestimmung von Kantenpunkten: Hier kommen sog. Kantenfilter zum Einsatz. Bei einem Filter in der Bildverarbeitung handelt es sich um einen Operator, der die Pixel eines Eingangsbilds über eine mathematische Abbildung neu ordnet und als Ausgangsbild ausgibt⁷⁰. Kantenfilter beispielsweise lokalisieren Bereiche im Bild, an denen starke Intensitätsänderungen vorkommen und liefern als Ergebnis ein entsprechendes Kantenbild. Die Neuberechnung eines Pixels erfolgt durch eine Matrix, welche in diesem Zusammenhang als Filtermaske bezeichnet wird. Verschiedene Filter liefern für unterschiedliche Eingabebilder unterschiedliche Ergebnisse, abhängig von den Eigenschaften ihrer jeweiligen Filtermaske. Zu den Kantendetektoren zählt der Sobel-Operator, welcher mathematisch relativ simpel ist, dessen Erkennungsgüte allerdings von der Verlaufsrichtung der Kanten abhängt, d.h. horizontal und verti-

70 Vgl. Jähne 2005, S. 111.

kal verlaufenden Linien werden gut erkannt⁷¹. Ein anderes Beispiel stellt der Laplace-Filter dar, dessen Ergebnis unabhängig von der Verlaufsrichtung der Kanten ist⁷². Weitere Verfahren sind u.a. der Prewitt-Operator und der Kirsch-Operator⁷³. Eines der wichtigsten Verfahren bei der robusten Identifikation von Kanten ist der sog. Canny-Algorithmus⁷⁴. Für dieses Verfahren definierte Canny folgende Kriterien⁷⁵:

- Es sollen keine Kanten übersehen werden und keine Kanten an Stellen identifiziert werden, an denen sich keine befinden.
- Kanten sollen möglichst genau dort lokalisiert werden, wo sie auch im Originalbild zu finden sind.
- Eine Kante soll nur an einer Stelle ausfindig gemacht werden, es soll nicht mehrere Antworten an mehreren Stellen geben.

Der Algorithmus lässt sich in drei Schritte einteilen. Im ersten Schritt erfolgt für das Bild eine Rauschunterdrückung (Glättung) über einen Gauß-Filter⁷⁶. Anschließend werden Betrag und Richtung der Gradienten einzelner Pixel ermittelt, dies kann z.B. geschehen über einen der bereits genannten Sobel- oder Prewitt-Operatoren. Daraus ergibt sich ein sog. Gradienten-Betragsbild. Hier erfolgt eine Unterdrückung derjenigen Pixel, welche entlang der Gradientenrichtung (d.h. senkrecht zur Kante) kein Maximum aufweisen⁷⁷. Auf diese Weise ergibt sich eine Kante mit der Breite von einem Pixel. Im dritten Schritt werden eine Ober- und Untergrenze für die Pixelwerte definiert. Die Obergrenze dient dazu, schwache Kanten zu unterdrücken, die z.B. aus Reustrauchen entstehen können⁷⁸. Mit der Untergrenze werden für die verbleibenden Kanten Unterbrechungen durch schwächere Kantenbereiche vermieden. Das Ergebnis ist ein sog. Kantenpunktbild.

- Bestimmung von Liniensegmenten: Die vorliegenden Linienpunkte werden zu Liniensegmenten zusammengefasst. Schreer unterscheidet dabei einerseits globale und andererseits sequentielle Segmentierungsverfahren⁷⁹. Zu den globalen Verfahren zählt die sog. Hough-Transformation, mit welcher geprüft werden kann, ob einzelne Segmente einer vorgegebenen geometrischen Referenzstruktur ähnlich sind, in diesem Falle einer Gerade⁸⁰. Für die Geradengleichung $ax+by+c=0$ gilt, dass die Gerade bei $b \neq 0$ durch die beiden Parameter Steigung

71 Vgl. Waldowski (Hg.) 2010, S 29.

72 Vgl. Steinmüller 2008, S. 33f.

73 Vgl. Neumann 2005, S. 59.

74 Vgl. Nischwitz; Fischer; Haberäcker 2007, S. 486.

75 Vgl. Canny 1986.

76 Vgl. Neumann 2005, S. 58.

77 Vgl. Nischwitz; Fischer; Haberäcker 2007, S. 488.

78 Vgl. Canny 1986.

79 Vgl. Schreer 2005, S. 158f.

80 Vgl. Nischwitz; Fischer; Haberäcker 2007, S. 506.

$m=-a/b$ und den y -Achsenabschnitt $t=-c/b$ ebenfalls eindeutig definiert ist. Darüber hinaus kann eine Gerade definiert werden, indem vom Koordinatenursprung das Lot auf die Gerade gefällt wird. Die Länge r des Lots und dessen Winkel mit der x - oder y -Achse φ legen die Gerade eindeutig fest. Der Zusammenhang zwischen (x,y) -Koordinaten und $(r;\varphi)$ -Koordinaten wird über die Hesse'sche Normalform hergestellt⁸¹:

$$r = x * \cos \varphi + y * \sin \varphi$$

Formel 8

In einem Koordinatensystem, dessen Achsen r (Abszisse) und φ (Ordinate) repräsentieren, entspricht der Geraden ein Punkt. Die Punkte des Kantenpunktbildes können in diesen neuen Parameterraum transformiert werden, wenn er diskretisiert wurde. Über die dort auftretende Lokalisierung können dann Segmente extrahiert werden.

Zu den sequentiellen Verfahren zählt beispielsweise die sog. Split-and-Merge-Technik. Dabei wird eine Kette von Punkten anhand eines Fehlerkriteriums solange unterteilt und ggf. wieder zusammengefasst, bis sich ein Minimum der Abweichungen aller Punkte von der geschätzten Geraden ergibt⁸². Bei einer anderen sequentiellen Methode werden Kantenpunkte verfolgt und zu Ketten zusammengefasst. Dabei wird von einem Startpunkt ausgegangen und in einer bestimmten Richtung nach weiteren Kantenpunkten gesucht. Dabei wird konstant geprüft, ob die gesuchten Punkte mit möglichst geringem Fehler durch ein Liniensegment angenähert werden können. Wird der Fehler für einen Punkt zu groß, bricht die die Segmentbildung ab und die noch verbleibenden Punkte werden untersucht. Um eine Übersegmentierung zu verhindern und zu gewährleisten, dass tatsächlich nur Objektkanten der Szenenobjekte erfasst werden, kann eine Mindestlänge für die Segmente definiert werden.

- Vorüberlegungen zur Korrespondenzanalyse: Um ein Liniensegment in Ansicht 1 seinem korrespondierenden Segment in Ansicht 2 korrekt zuzuordnen zu können, können verschiedene Eigenschaften genutzt werden. Ausgehend von ihrer zugehörigen Objektkante weisen einige Segmente andere Eigenschaften auf als andere, wodurch sie untereinander unterscheidbar sind. Intrinsische Eigenschaften beschreiben den Verlauf der Linie innerhalb des Bildbereichs. Dieser kann entweder über die Anfangs- und Endpunkte eines Liniensegments oder über den Mittelpunkt des Segments gemeinsam mit der Länge und der Orientierung angegeben werden⁸³. Da korrespondierende Linien eines Stereobildpaares aufgrund der perspektivischen Verzerrung teilweise sehr unterschiedliche Ver-

81 Vgl. Nischwitz; Fischer; Haberäcker 2007, S. 506.

82 Vgl. Schreer 2005, S. 159.

83 Vgl. Schreer 2005, S. 160.

läufe haben können, sind intrinsische Eigenschaften für die Korrespondenzanalyse nicht immer die optimale Wahl. Die Grauwerteigenschaften von Liniensegmenten sind nicht von der perspektivischen Verzerrung betroffen. Definiert man Merkmale, die auf der Intensität entlang eines Liniensegments beruhen, lässt sich für den Gradienten folgende Formel aufstellen⁸⁴:

$$Grad_{seg} = I_{l/o} - I_{r/u}$$

Formel 9

$I_{l/o/u}$ bezeichnet dabei die mittlere Intensität links und rechts, bzw. ober- und unterhalb des Segmentes (siehe Abb. 34).

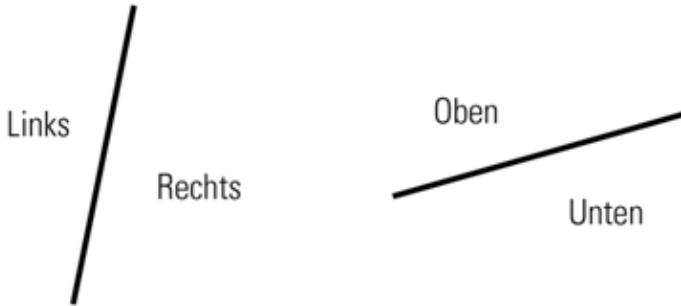


Abbildung 7: Segmentseiten

Die Richtung der Intensitätsänderung (DIR_{seg}) wird indiziert durch das Vorzeichen des Gradienten. Der mittlere Grauwert (MWG_{seg}) entlang der Kanten wird bestimmt durch den Mittelwert der Intensitäten auf den beiden Seiten des Liniensegments⁸⁵:

$$DIR_{seg} = \text{sign}(Grad_{seg})$$

Formel 10

$$MWG_{seg} = \frac{(I_{l/o} + I_{r/u})}{2}$$

Formel 11

Die mittlere Intensität beiderseits des Liniensegments wird über die Bildpunkte in einer bestimmten Region berechnet. Der Rand gibt den Abstand der Region zum Anfang und Ende des Segments an. Die Distanz bezeichnet den Abstand

⁸⁴ Vgl. Schreer 2005, S. 161.

⁸⁵ Vgl. Schreer 2005, S. 161.

der Region vom Segment über die gesamte Länge. Des Weiteren wird die Breite der Region definiert, die zur Berechnung der mittleren Intensität herangezogen wird. Der Mittelwert der Grauwerte des betrachteten Bildbereiches liefert die mittlere Intensität an jeder Seite des Segments:

$$I_{l/r/o/u} = \frac{1}{|\Lambda|} \sum_{i \in \Lambda(\text{Rand}, \text{Breite}, \text{Distanz})} x(i)$$

Formel 12

mit $|\Lambda|$ als Menge aller Bildpunkte der Region und $x(i)$ als Grauwert der Region Λ . Die Verwendung der Grauwerteigenschaften führt zu einer gesteigerten Zuverlässigkeit in der Zuordnung und senkt gleichzeitig den Rechenaufwand für die Korrespondenzanalyse⁸⁶.

Zusätzlich zu den segmenteigenen Eigenschaften können zur Korrespondenzanalyse die Beziehungen mehrerer benachbarter Segmente betrachtet werden. Zur Beschreibung der verschiedenen möglichen Kombinationen existieren in der Literatur mehrere Ansätze. Ein Vorschlag zur Beschreibung ergibt sich aus dem Schnittpunkt der Segmente und den Winkeln, die für die dadurch konstruierten Flächen entstehen. Es werden verschiedene Typen definiert, die sich anhand des Segmentverlaufs und der Winkelgrößen unterscheiden oder zuordnen lassen⁸⁷. Ein zweiter Ansatz besteht in einer Strukturdefinition, die abhängig von der Anzahl der Liniensegmente verschiedene Komplexitätsstufen beschreibt⁸⁸.

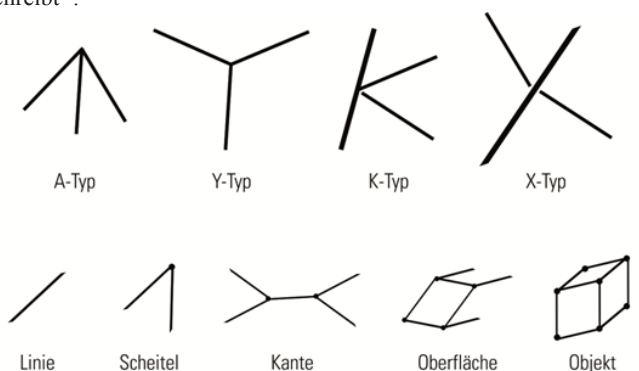


Abbildung 8: Segmentkombinationen

86 Vgl. Schreer; Hartmann; Adams 1997.

87 Vgl. Gu; Wu1990.

88 Vgl. Hourad; Skordas 1998.

• **Korrespondenzanalyse:** Die Korrespondenzanalyse für Liniensegmente lässt sich in die drei Phasen Prädiktion, Propagierung und Validierung gliedern. In einem Vorverarbeitungsschritt werden die Bilder beider Ansichten in Fensterbereiche eingeteilt, wodurch sich Nachbarschaftsbeziehungen zwischen Segmenten aufstellen lassen, welche im gleichen Fenster liegen⁸⁹. Bei Fenstern mit identischer Größe kann es vorkommen, dass zwischen bestimmten Segmenten, welche eigentlich benachbart sind, keine Beziehung hergestellt wird. Aus diesem Grund werden zusätzlich überlappende Teilfenster genutzt, die eine relative Verschiebung zum ersten Fenstermuster aufweisen. Auf diese Weise lassen sich für die Segmente Nachbarschaftsgraphen bestimmen⁹⁰.

• **Prädiktion:** In der ersten Phase werden für Segmente der linken Ansicht potenzielle Zuordnungen in der rechten Ansicht bestimmt. Hierfür wird zunächst die Epipolarbedingung genutzt. Ausgehend vom Mittelpunkt eines Segments in der linken Ansicht wird geprüft, ob die zugehörige Epipolarlinie in der rechten Ansicht von einem Segment geschnitten wird. Alle entsprechend betroffenen Segmente werden dann über ihre intrinsischen Eigenschaften Länge und Orientierung sowie über ihre Grauwerteigenschaften mit dem Segment der linken Ansicht verglichen⁹¹. Dieser Prozess wird für sämtliche Liniensegmente in der linken Ansicht durchgeführt. In der Regel lässt sich dadurch noch keine Eindeutigkeit herstellen, zumindest jedoch wird die Zahl der möglichen korrespondierenden Kandidaten verringert.

• **Propagierung:** Um eine globale Konsistenz für die Korrespondenzanalyse zu erreichen, wird die Nachbarschaft von Liniensegmenten innerhalb eines Bildes betrachtet, indem auf die Nachbarschaftsgraphen aus dem Vorverarbeitungsschritt zurückgegriffen wird. Ausgehend von einer Prädiktion für ein Segment des linken Bildes werden die Nachbarschaftsgraphen beider Ansichten rekursiv abgearbeitet, wobei als zusätzliches Kriterium die Glattheitsbedingung erfüllt werden muss⁹². Die Propagierung liefert dadurch für alle in der Prädiktion ermittelten Segmente der linken Ansicht mehrere sog. Hypothesen von korrespondierenden Segmenten. Die Länge einer Hypothese bezeichnet dabei die Anzahl der Segmente, die im Rahmen der rekursiven Abarbeitung alle Ähnlichkeitsbedingungen erfüllt haben. Innerhalb einer Hypothese kann ein Segment nur einmal auftreten, allerdings kann ein Segment in mehreren Hypothesen vorkommen. Um Eindeutigkeit herzustellen, muss eine Validierung erfolgen.

• **Validierung:** Hier besteht das erste Kriterium in der Länge der Hypothese, d.h. die Anzahl der korrespondierenden Segmente innerhalb einer räumlichen

89 Vgl. Schreer 2005, S. 164.

90 Vgl. Schreer 2005, S. 165.

91 Vgl. Schreer 2005, S. 166.

92 Vgl. Schreer 2005, S. 167.

Nachbarschaft. Je länger die Hypothese ist, desto mehr Ähnlichkeitsbedingungen wurden erfüllt und umso größer ist daher auch die Wahrscheinlichkeit, dass ein Segment, welches in einer längeren Hypothese enthalten ist, korrekt zugeordnet wurde⁹³. Bei Hypothesen gleicher Länge reicht dieses Kriterium nicht aus. In diesem Fall wird anhand der geometrischen und Grauwerteigenschaften ein bestes lokales Ähnlichkeitsmaß zwischen einem Segment in der linken und einem Segment in der rechten Ansicht bestimmt.

2.2.6 Tiefenrekonstruktion über Stereotriangulation

Die weiteren Betrachtungen referenzieren nun auf den vorangegangenen Artikel „Einführung in die Bildverarbeitung stereoskopischer Views“.

Aus der Stereoanalyse liegt die Disparität für korrespondierende Bildpunkte \mathbf{m}_l und \mathbf{m}_r vor. Mit Hilfe dieser Punkte kann eine Rekonstruktion des zugehörigen Objektpunktes \mathbf{M}_w erfolgen. Dies wird bewerkstelligt, indem die Koordinaten des Objektpunktes für ein euklidisches Koordinatensystem ermittelt werden. Dies ist dann möglich, wenn die perspektivischen Projektionsmatrizen der beiden Kameras bekannt sind. Im Allgemeinen ist davon auszugehen, dass in den erläuterten Verfahren die Punktkorrespondenzen nicht mit hundertprozentiger Genauigkeit hergestellt werden können. Dies bedeutet, dass die beiden optischen Strahlen, die durch die Bildpunkte \mathbf{m}_l und \mathbf{m}_r und die Brennpunkte C_l und C_r konstruiert werden, sich nicht an den exakten Koordinaten von \mathbf{M}_w schneiden⁹⁴. Für die tatsächliche Lage des Objektpunktes liefert der Mittelpunkt der kürzesten Strecke zwischen den beiden optischen Strahlen eine gute Annäherung. Mathematisch betrachtet handelt es sich hierbei um eine Optimierung im Sinne des kleinsten quadratischen Abstandes zwischen den optischen Strahlen⁹⁵. Um die notwendigen Gleichungen aufstellen zu können, wird auf die Formel der allgemeinen perspektivischen Projektionsmatrix zurückgegriffen:

$$s\tilde{\mathbf{m}}_i = \mathbf{P}_i \tilde{\mathbf{M}}_w \quad \text{mit } \mathbf{P}_i = \mathbf{A}_i \mathbf{P}_{Norm} \mathbf{D}_i, \mathbf{D}_i = \begin{bmatrix} \mathbf{R}_i & \mathbf{t}_i \\ \mathbf{0}_3^T & 1 \end{bmatrix} \quad \text{und } \mathbf{0}_3 = [0,0,0]^T$$

Formel 13

Wie vorab erläutert, enthält die Matrix \mathbf{A} die intrinsischen Parameter der jeweiligen Kamera, während über \mathbf{D}_i der Bezug zu einem gemeinsamen Weltkoordinatensystem hergestellt wird. Für den weiteren Verlauf bezeichnet p_{jk}^i das (j,k) -Element der Kamera i . Die ersten drei Elemente jeder Zeile von \mathbf{P}_i werden in Form eines Vektors $\mathbf{p}_i^j = (\mathbf{p}_{i1}^j, \mathbf{p}_{i2}^j, \mathbf{p}_{i3}^j)^T$ zusammengefasst. Unter Nutzung des Vektorprodukts entfällt der homogene Skalierungsvektor⁹⁶:

93 Vgl. Schreer 2005, S. 168.

94 Vgl. Hartley; Zisserman 2003, S. 312.

95 Vgl. Schreer 2005, S. 176.

96 Vgl. Hartley; Zisserman 2003, S. 312.

$$\tilde{\mathbf{m}}_i \times \mathbf{P}_i \tilde{\mathbf{M}}_w = 0 \text{ bzw.}$$

Formel 14

$$v_i(\mathbf{p}_3^i \tilde{\mathbf{M}}_w) - (\mathbf{p}_2^i \tilde{\mathbf{M}}_w) = 0$$

Formel 15

$$u_i(\mathbf{p}_3^i \tilde{\mathbf{M}}_w) - (\mathbf{p}_1^i \tilde{\mathbf{M}}_w) = 0$$

Formel 16

$$u_i(\mathbf{p}_2^i \tilde{\mathbf{M}}_w) - v_i(\mathbf{p}_1^i \tilde{\mathbf{M}}_w) = 0$$

Formel 17

• Homogene Methode: Über die beiden ersten linear unabhängigen Gleichungen für die Werte \mathbf{u} und \mathbf{v} kann beim Vorhandensein von zwei Punktkorrespondenzen folgende Gleichung aufgestellt werden⁹⁷:

$$\mathbf{B} \tilde{\mathbf{M}}_w = 0 \quad \text{mit } \mathbf{B} = \begin{bmatrix} u_1 \mathbf{p}_3^{1T} - \mathbf{p}_1^{1T} \\ v_1 \mathbf{p}_3^{1T} - \mathbf{p}_2^{1T} \\ u_2 \mathbf{p}_3^{2T} - \mathbf{p}_1^{2T} \\ v_2 \mathbf{p}_3^{2T} - \mathbf{p}_2^{2T} \end{bmatrix}$$

Formel 18

Die Gleichung kann mit Hilfe der Direkten Linearen Transformation (kurz: DLT) durch Lösen des Eigenwertproblems gelöst werden⁹⁸. Als Nebenbedingung wird dabei $\|\tilde{\mathbf{M}}_w\| = I$ verwendet.

Inhomogene Methode: Wird die vierte Komponente von $\tilde{\mathbf{M}}_w$ auf Eins gesetzt, ergeben sich vier inhomogene Gleichungen mit den Unbekannten X_w, Y_w und Z_w . Für die korrespondierenden Punkte \mathbf{m}_i ergibt sich:

$$(\mathbf{p}_1^1 - u_1 * \mathbf{p}_3^1)^T * \mathbf{M}_w + p_{14}^1 - u_1 * p_{34}^1 = 0$$

Formel 19

⁹⁷ Vgl. Hartley; Zisserman 2003, S. 312.

⁹⁸ Vgl. Schreier 2005, S. 177.

$$(\mathbf{p}_2^1 - v_1 * \mathbf{p}_3^1)^T * M_w + p_{24}^1 - v_1 * p_{34}^1 = 0$$

Formel 19

$$(\mathbf{p}_1^2 - u_2 * \mathbf{p}_3^2)^T * M_w + p_{14}^2 - u_1 * p_{34}^2 = 0$$

Formel 20

$$(\mathbf{p}_2^2 - v_2 * \mathbf{p}_3^2)^T * M_w + p_{24}^2 - v_2 * p_{34}^2 = 0$$

Formel 21

Bei $\mathbf{B}_i = \begin{pmatrix} (\mathbf{p}_1^i - u_i * \mathbf{p}_3^i)^T \\ (\mathbf{p}_2^i - v_i * \mathbf{p}_3^i)^T \end{pmatrix}$ und $\mathbf{c}_i = \begin{pmatrix} u_i * p_{34}^i - p_{14}^i \\ v_i * p_{34}^i - p_{24}^i \end{pmatrix}$ für $i = 1, 2$ ergibt

sich für beide Kameras:

$$\mathbf{B} * M_w = \mathbf{c} \text{ mit } \mathbf{B} = \begin{pmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{pmatrix}, M_w = (X_w, Y_w, Z_w)^T \text{ und } \mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

Formel 22

Unter der Voraussetzung, dass $\mathbf{B}^T \mathbf{B}$ invertierbar ist, ergibt sich für den Objektpunkt über das Problem der Minimierung durch den kleinsten quadratischen Fehler folgende Formel:

$$M_w = (\mathbf{B}^T * \mathbf{B})^{-1} * \mathbf{B}^T * \mathbf{c}$$

Formel 23

2.3 Zusammenfassung

Zur Gewinnung von Raumtiefe existieren verschiedene Verfahren. Für ein Stereokamerasystem kann die Triangulation sehr gut genutzt werden; hierzu müssen zuvor in den beiden Stereoansichten Korrespondenzen zwischen Bildmerkmalen wie Punkten oder Linien bestimmt werden. Für diese Stereoanalyse gelten bestimmte Bedingungen, welche in den unterschiedlichen Methoden zur Ermittlung von Korrespondenzen genutzt werden können. Liegen die Korrespondenzen vor, so können aus diesen zusammengehörenden Bildpunkten die Koordinaten ihrer jeweiligen Szenenpunkte errechnet und damit die räumliche Tiefe der Szene bestimmt werden.

3 Tests

Im Rahmen der Untersuchung sollten stereoskopische Testaufnahmen durchgeführt werden, um praktische Kenntnisse hinsichtlich der Erstellung von Tiefenkarten über Stereotriangulation zu gewinnen und die Kenntnisse des bisher Dargelegten anzuwenden, bzw. diese – sofern möglich – zu bestätigen.

3.1 Planung

3.1.1 Kameras

Als Startkonfiguration wurden zwei einfache Camcorder vom Typ NV-GS500 der Firma Panasonic gewählt. Beiden Geräte wurden zur Aufnahme auf der Stereoschiene, welche Anja Frank im Rahmen ihrer Bachelorthesis⁹⁹ im Sommersemester 2007 in der Werkstatt der HFU hatte anfertigen lassen, befestigt. Für die Ausrichtung der optischen Achsen wurde die achsparallele Version gewählt, da keine befriedigende Möglichkeit bestand, die Winkel einer konvergierenden Anordnung hinreichend exakt zu messen.



Abbildung 9: Verwendetes Stereorig

3.1.2 Software

Als Programm zur Erzeugung der Tiefenkarten wurde eine Software der russischen Firma Triaxes Ltd. mit dem Namen Stereotracer eingesetzt. Die Firma bietet mehrere

⁹⁹ Vgl. Frank 2007.

Softwarepakete an, deren Hauptzweck darin besteht, Bilder so aufzubereiten, dass diese mit Hilfe der Lentikulartechnik auf geeignete Träger gedruckt werden können. Die Stereotracer-Software ist kostenpflichtig, es besteht jedoch die Möglichkeit, sie in einem Demomodus auch kostenlos zu betreiben. In diesem Modus ist der volle Funktionsumfang verfügbar, die Bilder werden allerdings mit einem Wasserzeichen exportiert. Die Hauptfunktion des Programms besteht in der Generierung von mehreren Ansichten einer Szene, wobei als Ausgangsmaterial entweder ein einzelnes Bild und eine zugehörige, bereits künstlich erzeugte Tiefenkarte oder zwei Stereoansichten dienen. Eine weitere Funktion der Software ermöglicht die Generierung von Tiefenkarten aus zwei Stereoansichten. Innerhalb der Anwendung können lediglich Einzelbilder verarbeitet werden. Stereotracer unterstützt dabei gängige Bildformate, u.a. Windows Bitmap (.bmp), das JPEG-Format (.jpg), Portable Networks Graphics (.png) und das Tagged Image File Format TIFF (.tif). Allerdings verfügt das Programm über eine Batch-Funktion, mit der mehrere Einzelbilder gleichzeitig importiert und deren Tiefenkarten automatisch generiert werden können. Diese Funktion war einer der maßgeblichen Gründe dafür, die Tiefenkarten mit Hilfe von Stereotracer zu generieren. Dadurch ließen sich auch komplette Sequenzen erstellen, sofern die Quelldaten in einem entsprechenden Format vorlagen.

3.1.3 Szenenauswahl und Arrangement

Mit den Testaufnahmen sollten möglichst viele Situationen erfasst werden, die während realer Dreharbeiten u.U. auftreten können. Es sollte festgestellt werden, wie gut die Tiefe bestimmter Bildbereiche unter den gegebenen Umständen erfasst werden kann. Dementsprechend wurde jede Szene hinsichtlich unterschiedlicher Aspekte ausgewählt, bzw. arrangiert. Es sollten u.a. folgende Situationen abgedeckt werden:

- Große, über einen weiten Verlauf einheitlich texturierte Flächen
- Schwache Beleuchtung
- Viele Verdeckungen
- viele Tiefensprünge (hier wurden zusätzlich eine Zu- und Wegfahrt durchgeführt, um die Auswirkungen der sich ändernden Verdeckungsverhältnisse zu beobachten)
- Außenaufnahmen mit großer räumlicher Tiefe
- Feine Details und Strukturen
- Einheitlicher Hintergrund mit wenig Tiefe und einem länglichen Objekt, welches in die Tiefe ragt. Hierbei zwei zusätzliche Situationen:
 - Tiefenposition des Objekts wird verändert
 - Unterschiedliche Beleuchtungssituationen mit besonderem Augenmerk auf den Auswirkungen des veränderten Schattenwurfs

3.2 Ablauf

Nachdem die Versuchs-Szene vorbereitet worden war, wurden verschiedene Werte festgehalten und über eine Checkliste dokumentiert (s. Anhang). Es handelte sich dabei für jede Szene um folgende Angaben:

- Höhe des optischen Mittelpunkts der Kamera (mm)
- Stereobasis (mm)
- Brennweite (mm)
- Verschlusszeit (sec)
- Blendenzahl
- Scheinfensterweite (mm)
- Nahpunktweite (mm)
- Fernpunktweite, sofern möglich (mm)

Zusätzlich wurde in jeder Szene ein Referenzobjekt platziert oder bestimmt, für das die räumlichen Abmessungen und der Nahpunkt ebenfalls notiert wurden.

- Nahpunktweite Referenzobjekt (mm)
- Abmessungen Referenzobjekt (mm)

3.3 Nachbearbeitung

3.3.1 Schnitt

Nach der Aufnahme wurden die Sequenzen auf eine Workstation überspielt und mit Adobe Premiere Pro eine grobe Schnittfassung erstellt. Der wichtigste Teil hierbei war, die Synchronität für linke und rechte Videospur aufrechtzuerhalten. Nachdem die Sequenzen auf die gewünschten Längen geschnitten waren, wurden sie als AVI-Video exportiert.

3.3.2 Manuelle Korrekturen

Mit Adobe After Effects wurde für die AVI-Sequenzen eine manuelle Disparitätskorrektur durchgeführt. Hierzu wurden beiden Spuren in der Zeitleiste übereinander platziert und die Transparenz der oberen Spur um 50 % reduziert. Auf diese Weise war die untere Spur durch die obere hindurch sichtbar und konnte an diese angeglichen werden. Für beide Ansichten wurde zunächst – sofern vorhanden – der vertikale Versatz korrigiert, indem sie in ihrem Höhenwert gleichmäßig zueinander verschoben wurden. Anschließend wurde über eine gegenseitige seitliche Verschiebung die horizontale Disparität durchgeführt. Dadurch, dass die Ansichten verschoben worden waren, entstanden an den Bildrändern schwarze Lücken. Diese wurden eliminiert, indem eine minimale Aufskalierung (2-5 %) beider Ansichten erfolgte.

In einem weiteren Arbeitsschritt wurden für einen Teil der Stereosequenzen Hellig-

keits- und Farbunterschiede behoben, welche bei der Aufnahme nicht zu identifizieren gewesen waren. Anschließend wurden beide Spuren separat als Quicktime-Video exportiert, um die vorgenommenen Korrekturen begutachten zu können. Adobe After Effects bietet eine stereoskopische, anaglyphe Vorschau ohne vorherigen Export. Der entsprechende Effekt „3D-Brille“ ignoriert allerdings horizontale und vertikale Positionsänderungen, die auf das Footage angewandt wurden. Aus diesem Grund wurde die o.g. Variante genutzt, welche zwar mehr Zeit in Anspruch nimmt, dafür allerdings die Disparitätskorrekturen berücksichtigt. Die exportierten Quicktime-Sequenzen wurden mit Hilfe der Player-Software Stereoscopic Player der Firma 3dtv.at in anaglypher Darstellung beurteilt. Gegebenenfalls wurden weitere Änderungen innerhalb von After Effects durchgeführt.

Waren die Ergebnisse befriedigend, wurden von beiden Spuren repräsentative einzelne Frames separat im TIFF-Format exportiert, um mittels Stereotracer eingelesen werden zu können. Die einzelnen Frames wiesen eine räumliche Auflösung von 1024×576 Pixel auf mit $PAR = 1$ und einem Seitenverhältnis von $16 : 9$.

3.3.3 Tiefenkarten für Einzelbilder

Beim Erstellen eines neuen Projekts in Stereotracer muss zuerst angegeben werden, welches Quellmaterial importiert werden soll. Wie bereits dargelegt, kann hierbei zwischen den Optionen 2D-Ansicht + Tiefenkarte oder linke Ansicht + rechte Ansicht gewählt werden. Es müssen dann jeweils die beiden entsprechenden Bilddateien ausgewählt werden. Mit Hilfe zweier Schieberegler kann dem Programm der Bereich der parallaxtischen Verschiebung angegeben werden. Anschließend werden mehrere vorläufige Vorschläge von Tiefenkarten präsentiert, die manuell angepasst werden können. Im Anschluss daran wird die finale Tiefenkarte erstellt und kann als Bilddatei exportiert werden.



Abbildung 10: Tiefenkarte aus Stereotracer

3.3.4 Tiefenkartensequenzen

Zur Verarbeitung einer Bildfolge über die Batch-Funktion von Stereotracer müssen die Quellbilder im Side-by-Side-Format vorliegen. Beide Ansichten befinden sich also nebeneinander angeordnet in der gleichen Bilddatei. Aus diesem Grund wurden die korrigierten Kompositionen für linke und rechte Ansicht in After Effects in einer neuen, gemeinsamen Komposition mit der doppelten Pixelanzahl in horizontaler Richtung platziert. So ergab sich eine Side-by-Side Anordnung. Die Komposition wurde abschließend als TIFF-Einzelbildsequenz exportiert. Somit lagen die Stereoansichten in einer gemeinsamen Bildsequenz in Side-by-Side-Anordnung vor, bei einer räumlichen Auflösung von 2048×576 Pixel.

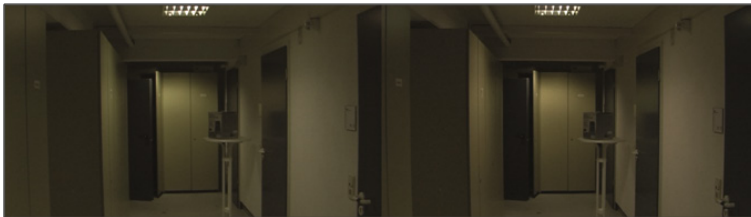


Abbildung 11: Side-by-Side-Anordnung

Beim Import in Stereotracer muss ein Zielordner festgelegt werden, in welchem die Tiefenkarten nach ihrer Erstellung gespeichert werden. Danach arbeitet das Programm automatisch alle ausgewählten Side-by-Side-Bilder ab. Als Resultat ergeben sich einzelne Bilddateien mit der gleichen Auflösung wie die Quelldatei. Dabei befindet sich die Tiefenkarte in der rechten Hälfte des Bildbereichs und die rechte Originalansicht in der linken Hälfte.



Abbildung 12: Tiefenkarte aus Side-by-Side

Um daraus eine Sequenz zu erhalten, welche ausschließlich nur die Tiefenkarten enthält, können die erstellten Bilder als Einzelbildsequenz erneut in After Effects importiert werden. Dort wird die Sequenz in einer Komposition platziert, welche die räumliche Auflösung einer der ursprünglichen linken oder rechten Ansicht aufweist, in diesem Fall also 1024×576 Pixel. Dabei muss das Footage so platziert werden, dass ausschließlich die linke Bildhälfte sichtbar ist, also genau die, in welcher die

Tiefenkarte enthalten ist. Mit diesen Einstellungen kann dann eine Sequenz im gewünschten Format exportiert werden, z.B. ebenfalls als Quicktime-Video.

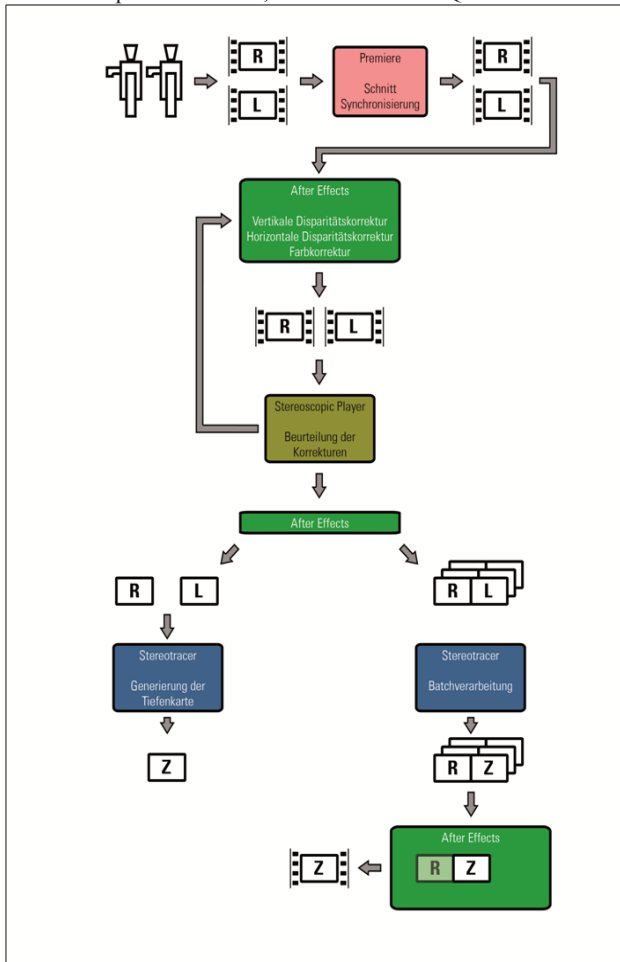


Abbildung 13: Stereoverarbeitungskette

3.3.5 Ausgewählte Ergebnisse

Im Folgenden werden für jede Szene ausgewählte repräsentative Stereobildpaare dargestellt.

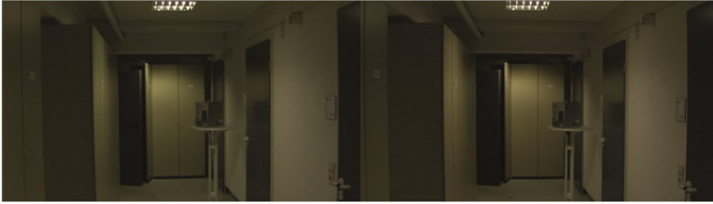


Abbildung 14: Tiefenkarte Keller 01



Abbildung 15: Tiefenkarte Schnittraum 01



Abbildung 16: Tiefenkarte Schnittraum 02_Fahrt



Abbildung 17: Tiefenkarte Schulhof 01



Abbildung 18: Tiefenkarte Wand 01

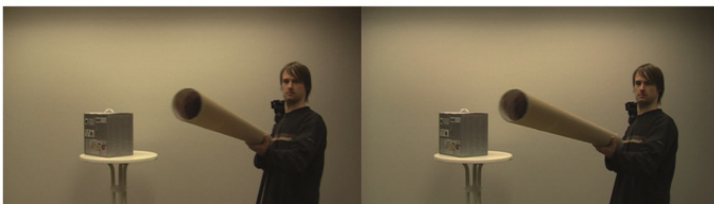


Abbildung 19: Tiefenkarte Wand 02_1

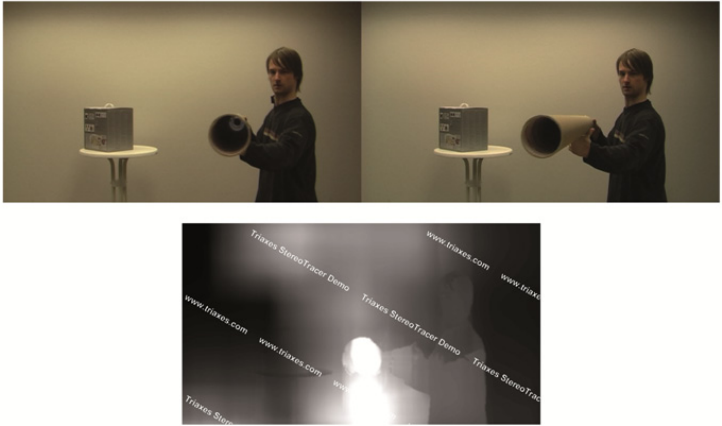


Abbildung 20: Tiefenkarte Wand 02_2

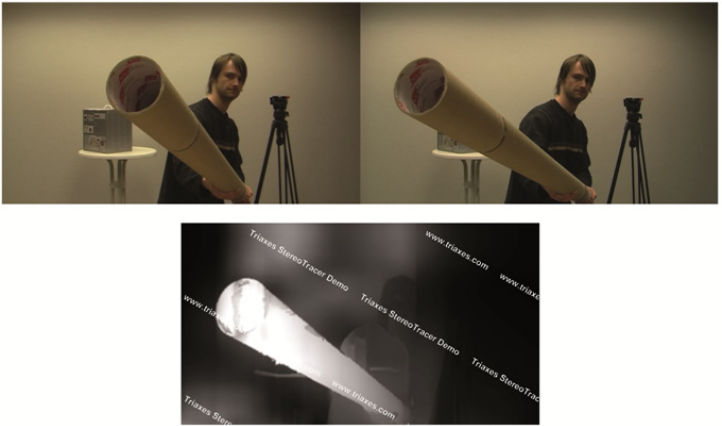


Abbildung 21: Tiefenkarte Wand 02_3

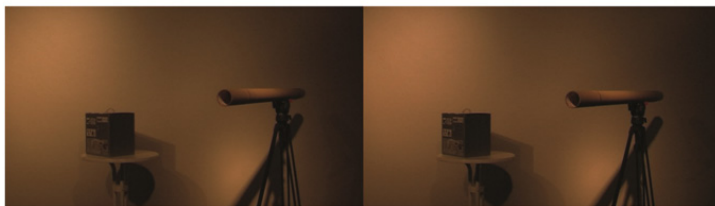


Abbildung 22: Tiefenkarte Wand 03_1

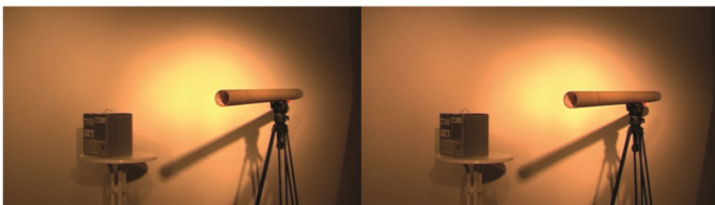


Abbildung 23: Tiefenkarte Wand 03_2



Abbildung 24: Tiefenkarte Wand 03_3

3.3.6 Analyse und Bewertung

In diesem Abschnitt erfolgt eine Analyse und Interpretation der auffälligsten Bereiche der Tiefenkarten.

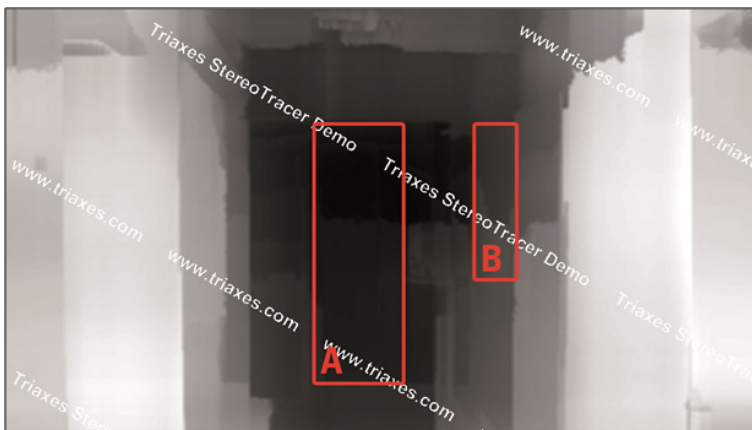


Abbildung 25: TK Analyse Keller 01

- A: Dem Schrank werden zwei unterschiedliche Tiefenebenen zugerechnet. Mögliche Erklärung: Helligkeitsunterschiede, verursacht durch die Beleuchtung

- B: Auch der Fläche neben der Schachtel als Referenzobjekt werden unterschiedliche Tiefenebenen zugewiesen. Mögliche Erklärung: Helligkeitsunterschiede, verursacht durch die Beleuchtung
- Insgesamt sind die Tiefenwerte relativ zuverlässig verteilt. Mögliche Erklärung: Das Vorhandensein vieler Kanten und damit Hell-Dunkel-Übergängen.

- Schnittraum 01



Abbildung 26: TK Analyse Schnittraum 01

- A: Die Tiefe der Lücke unter dem Tisch wird sehr gut erfasst
- B: Bei der Kante, die horizontal an der Wand verläuft, gibt es keine Tiefensprünge.
- C: Die Trennung der unterschiedlichen Objekte auf dem Tisch erfolgt relativ genau
- D: Die linke obere Tischkante geht direkt über in die Tiefenebene des Schrankes. Mögliche Erklärung: Die Kante wirkt in diesem Fall als Überleitung, da sie gleichzeitig Ähnlichkeit mit dem Farbton der Tischplatte und dem Farbton des Schrankes hat.
- E: Bildschirmröhre und Plastikgehäuse des Computermonitors weisen fälschlicherweise starke Tiefenunterschiede auf. Mögliche Erklärung: Starker Hell-Dunkel-Übergang zwischen den Bereichen.
- F: Für die Schachteln auf dem Schrank werden falsche Zuordnungen getroffen

fen, sie werden als tiefer liegend interpretiert als der Schrank selbst. Mögliche Erklärung: Die Kante zwischen den Elementen wird als Sprung in der Tiefe interpretiert, in diesem Fall wird der obere Bereich fälschlicherweise als tiefer positioniert dargestellt.

- G: Es ist kein Unterschied zwischen dem Boden und der Schrankfläche erkennbar, obwohl ein markanter Übergang vorhanden ist. Mögliche Erklärung: Die Flächen sind sehr einheitlich, da sie kaum markante Muster aufweisen und ähnliche Farbtöne besitzen.

- H: Die Tiefensprünge bei den Tisch- und Stuhlbeinen unterhalb der dahinter verlaufenden Schrankkante werden relativ ungenau erkannt. Mögliche Erklärung: Innerhalb des Bereichs herrscht ein ähnlicher Farbverlauf.

- I: Den unterschiedlichen Bestandteilen im oberen Bereich des Bürostuhls wird eine einheitliche Tiefe zugeschrieben. Mögliche Erklärung: Es gibt kaum Hell-Dunkel-Übergänge, die Farbtöne sind untereinander sehr ähnlich.

- **Schnittraum 02 Fahrt**

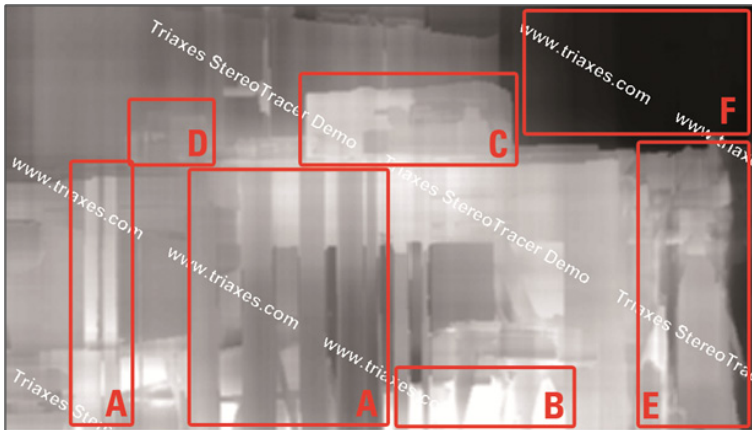


Abbildung 27: TK Analyse Schnittraum 02_Fahrt

- A: Die Tischbeine in der Mitte werden korrekt als separate Bereiche erkannt, allerdings wird ein falscher Tiefenwert zugewiesen. Dadurch werden die Bereiche fälschlicherweise über ihren eigentlichen Verlauf hinaus erweitert. Mögliche Erklärung: Die Beine gehen an einem Bereich in den Tisch über, der als einheitliche Fläche wirkt, da er ähnlichen Farbtönen und eine einheitliche Texturierung aufweist. Das gleiche Problem ist bei den beiden Tischbeinen im linken Bildbereich zu beobachten.

- B: Dem Bodenbereich, welcher hinter den Stuhlbeinen liegt, wird eine falsche

Tiefe zugewiesen. Mögliche Erklärung: Die durch die Verdeckungen verursachten Tiefsprünge werden genau umgekehrt interpretiert, da die Flächen in unterschiedlichen Tiefen für sich relativ einheitlich wirken. Die schwarzen Tisch- und Stuhlbeine könnten ebenso Lücken in einer Vordergrundfläche darstellen.

- C: Die Schachteln werden gut erfasst, allerdings nicht sehr gut voneinander getrennt. Mögliche Erklärung: Zur Umgebung hin verlaufen deutliche Kanten, untereinander weisen die Objekte sehr ähnliche Texturen auf.
- D: Die linke Schachtel wird schlecht vom Hintergrund getrennt, eine Erklärung hierfür fällt schwer.
- E: Die Tiefenzuweisungen für die Jacke über dem Stuhl sind sehr unterschiedlich. Mögliche Erklärung: Das Objekt besitzt relativ gut voneinander abgegrenzte Flächen, die untereinander allerdings lediglich geringe Tiefenunterschiede aufweisen.
- F: In diesem Bereich ist die Zuweisung korrekt, hier herrscht die größte Tiefe.

• Schulhof 01

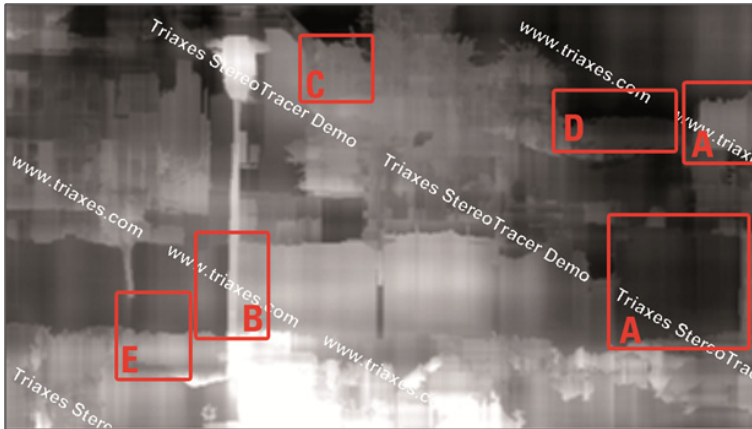


Abbildung 28: TK Analyse Schulhof 01

- A: Für den Schnee wird fälschlicherweise eine größere Tiefe ermittelt als für den Wald am Horizont. Mögliche Erklärung: Einheitliche und kontrastarme Flächen.
- B: Für die Bereiche links und rechts des Laternenpfahls verursachen Kanten Tiefsprünge, obwohl keine vorhanden sind. Mögliche Erklärung: Einheitliche und kontrastarme Fläche.

- C: In diesem Bereich werden Tiefensprünge nicht erfasst, der Hauswand wird die gleiche Tiefe zugewiesen wie dem Himmel. Mögliche Erklärung: Die Äste, die in den Bereich ragen, verhindern eine Erfassung der unterschiedlichen Tiefenebenen.
- D: Den weißen Wolken wird eine geringere Tiefe zugesprochen als dem blauen Himmel. Physikalisch betrachtet ist dies zwar korrekt, für die Tiefenkarte macht die Zuordnung wenig Sinn, da die Tiefenwerte beider Objekte gegen Unendlich gehen müssten. Mögliche Erklärung: Unterschiedliche Farbtöne.
- E: Der Schnee auf der Hecke wird korrekt vom Schnee hinter der Hecke getrennt, trotz eines scheinbaren Überganges. Mögliche Erklärung: Der Kantenverlauf der Hecke wird über die Unterbrechung hinweg fortgeführt.

• Wand 01

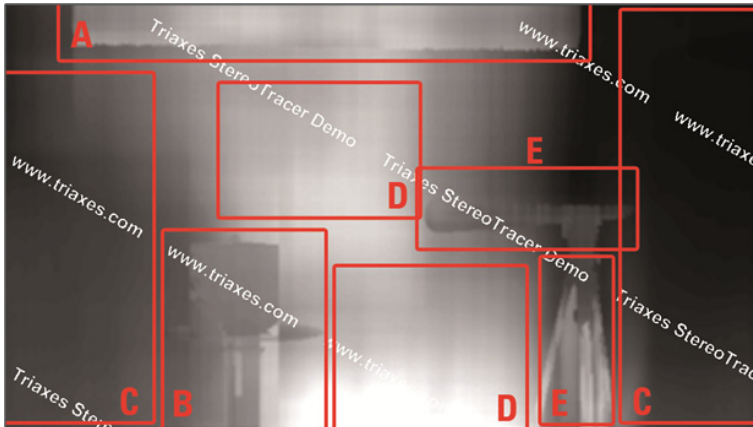


Abbildung 29: TK Analyse Wand 01

- A: Der obere Bildbereich wird als eigenständige Fläche interpretiert, obwohl er die gleiche Tiefe wie die restliche Wand haben müsste. Mögliche Erklärung: Durch die Beleuchtung entstehen Unterschiede im Helligkeitsverlauf.
- B: Der Tisch und die Schachtel werden relativ gut erkannt. Die zugewiesene Tiefe ist im Grunde korrekt, durch die falsche Zuweisung der Bereiche in der Bildmitte wirkt sie allerdings auf den ersten Blick falsch.
- C: Die Bereiche am linken und rechten Bildrand sind ziemlich genau zugeordnet, hier besteht die größte Tiefe.
- D: In diesen Bereichen ist die Tiefe falsch zugeordnet. Mögliche Erklärung: Beleuchtungsunterschiede führen zu unterschiedlichen Helligkeitsverläufen.

- E: Die Pappröhre wird gut erfasst, allerdings wird ihre Tiefe nicht vollständig korrekt zugeordnet. Der Tiefenwert ändert sich über den Verlauf hin kaum, obwohl er dies bei korrekter Erkennung eigentlich sollte. Mögliche Erklärung: An der oberen Kante der Röhre herrscht ein dem Hintergrund ähnlicher Farbton.
- F: Die Struktur des Statives wird gut erfasst, auch die Tiefenwerte werden relativ genau zugewiesen. Mögliche Erklärung: Es treten auffällige Hell-Dunkel-Übergänge zwischen den Stativbeinen und dem Hintergrund auf.

• Wand 02_1

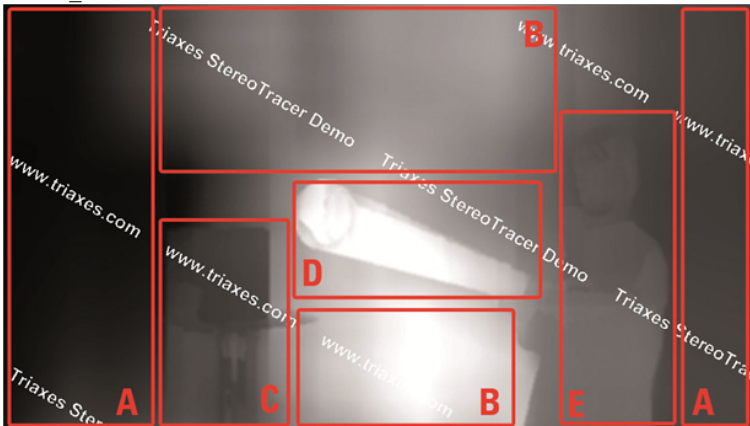


Abbildung 30: TK Analyse Wand 02_1

- A: Der linke und rechte Bildbereich wird korrekterweise je als einheitliche Tiefenebene erfasst. Allerdings werden den beiden Teilen unterschiedliche Tiefenwerte zugeordnet, obwohl sie gleich sein müssten. Mögliche Erklärung: Die Beleuchtung führt zu unterschiedlichen Helligkeiten.
- B: Diesem Bereich wird eine falsche Tiefe zugeordnet, sie müsste die gleichen Werte wie der linke und rechte Bildbereich aufweisen. Das Gleiche gilt in noch stärkerem Maß für den mittleren Bereich am unteren Bildrand. Mögliche Erklärung: Unterschiedliche Beleuchtungseigenschaften und schwach texturierte Flächen.
- C: Die Schachtel und der Tisch werden relativ gut segmentiert, auch die Tiefenwerte sind korrekt. Es werden jedoch erneut die Kantenverläufe der Schachtel fälschlicherweise über ihren eigentlichen Verlauf hinaus erweitert. Mögliche Erklärung: Die Unterbrechung durch die Tischfläche wird aufgrund der ähnlichen Farbtöne von Schachtel und Hintergrund ignoriert.

- D: Die Pappröhre wird ziemlich genau erfasst und auch der Verlauf der Tiefenwerte entspricht in etwa dem tatsächlichen Verlauf.
- E: Die Umrisse der Person werden ebenfalls gut erfasst, auch die Tiefe wird korrekt zugeordnet. Mögliche Erklärung: Ausgeprägte Hell-Dunkel-Übergänge zwischen Objekt und Hintergrund.

• Wand 02_2

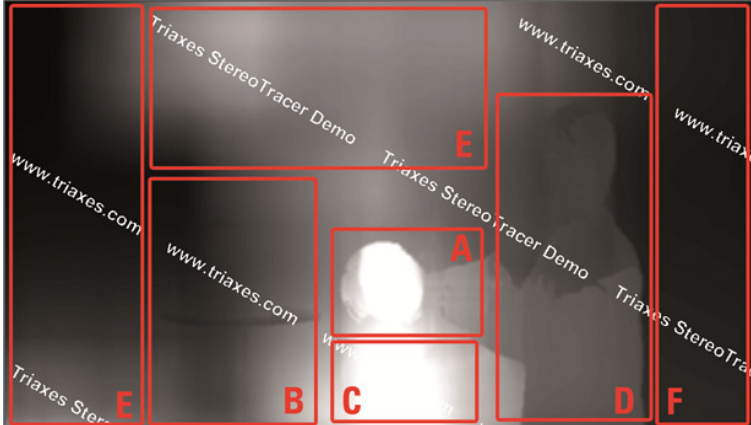


Abbildung 31: TK Analyse Wand 02_1

- A: Von der Pappröhre wird lediglich der Wert für den vorderen Bereich korrekt zugeordnet, vom Verlauf in die Tiefe ist nichts zu erkennen. Mögliche Erklärung: Aufgrund des perspektivischen Unterschiedes wird die Röhre in der linken Ansicht lediglich als Kreis dargestellt, somit können keine Zuordnungen für das rechte Stereohalbbild durchgeführt werden und die Tiefeninformation geht verloren.
- B: Dem Tisch und der Schachtel wird die gleiche Tiefe zugewiesen wie dem Hintergrund, obwohl kein Unterschied zur Konstellation von Bild Wand 2.1 besteht. Dieses Phänomen bleibt rätselhaft.
- C: Die Tiefe dieser Region ist falsch zugeordnet. Bei richtiger Zuordnung wäre dieser Bereich einer der dunkelsten. Auch diese Zuordnung ist schwer nachzuvollziehen.
- D: Die Person wird relativ gut erfasst. Es sind sogar Tiefenunterschiede innerhalb der Silhouette erkennbar. Mögliche Erklärung: Durch die Beleuchtung ergeben sich Helligkeitsunterschiede, die die Tiefe betonen.
- E: Der Hintergrund weist unterschiedliche Tiefenwerte auf. Mögliche Erklärung:

• F: Die Tiefe für den Hintergrund im rechten Bildbereich ist recht genau zugeordnet.

• Wand 02_3

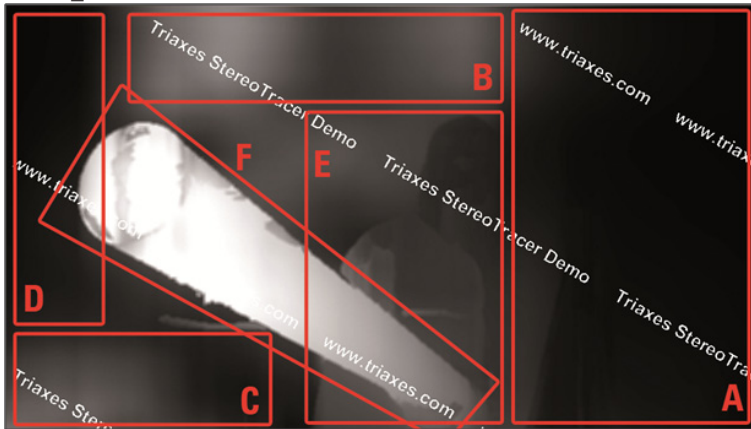


Abbildung 32: TK Analyse Wand 02_3

- A: Die Tiefe des rechten Bildbereichs ist korrekt erfasst
- B: Dem oberen Teil des Hintergrundes ist eine falsche Tiefe zugewiesen. Mögliche Erklärung: Die Beleuchtung führt zu Helligkeitsunterschieden.
- C: Dem Tisch wird fälschlicherweise die gleiche Tiefe zugewiesen wie dem Hintergrund. Mögliche Erklärung: Ein erheblicher Teil des Tisches und der Schachtel sind durch den perspektivischen Unterschied in der rechten Ansicht verdeckt. Somit ist es schwer, korrekte Korrespondenzen herzustellen.
- D: Die Tiefe für den Hintergrund links oben ist korrekt zugeordnet
- E: Der Tiefenverlauf der Person ist gut erfasst.
- F: Der Tiefenverlauf der Papprolle ist gut erfasst.

- Wand 03_1

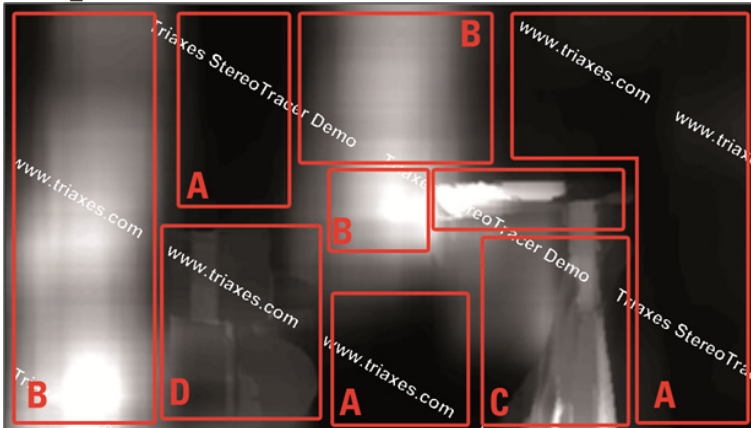


Abbildung 33: TK Analyse Wand 03_1

- A: In diesen Bildbereichen ist die Tiefe korrekt erfasst.
- B: Die Tiefe in diesem Bereich ist falsch zugeordnet, bei einer korrekten Zuordnung wäre der Bereich wesentlich dunkler. Mögliche Erklärung: Die Fläche besitzt eine sehr einheitliche Texturierung und wenige Unterschiede in den Kontrastabstufungen.
- C: Die Struktur des Statives und seiner Tiefenwerte werden relativ genau erfasst. Es erfolgt jedoch keine Abgrenzung zum Hintergrund im Bereich der Stativbeine. Mögliche Erklärung: Der Schatten lässt den gesamten Bereich in einem einheitlichen Farbton erscheinen.
- D: Der Tisch und die Schachtel werden vergleichsweise gut erfasst. Es treten allerdings Probleme aufgrund des Schattenwurfs auf. So wird dem Schatten der Tischplatte fälschlicherweise eine separate Tiefenebene zugewiesen.
- E: Die Pappröhre wird in ihrem Tiefenverlauf gut erfasst. Allerdings treten abrupte Tiefenänderungen auf, so dass der Röhre insgesamt drei unterschiedliche Tiefenebenen zugewiesen werden. Mögliche Erklärung: Durch die Beleuchtung erscheint der mittlere Bereich der Röhre heller und der hintere Bereich sehr dunkel.

- Wand 03_2

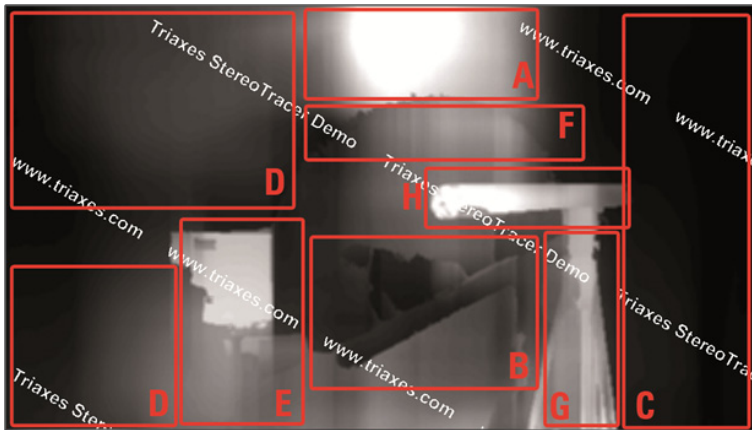


Abbildung 34: TK Analyse Wand 03_2

- A: Dem Bereich wird eine falsche Tiefe zugewiesen. Mögliche Erklärung: Es herrscht ein starker Helligkeitsunterschied zum unteren Bildbereich.
 - B: Hier wird der Hintergrund als eine Struktur mit unterschiedlichen Tiefenebenen interpretiert. Mögliche Erklärung: Der Schattenwurf der Pappröhre erzeugt vermeintliche Flächen.
 - C: Hier ist die Tiefe korrekt zugewiesen.
 - D: Der Verlauf der unterschiedlichen Tiefenwerte entspricht nicht dem tatsächlichen Bildinhalt. Mögliche Erklärung: Die Beleuchtung verursacht einen Helligkeitsverlauf, welcher falsch interpretiert wird.
 - E: Für einen Teil der Schachtel wird die Struktur gut erfasst und auch die richtige Tiefe zugewiesen. Allerdings gibt es Probleme auf dem Tisch, Schachtel und Hintergrund ineinander übergehen. Mögliche Erklärung: Die verschiedenen Elemente weisen durch die Beleuchtung ähnliche Farbwerte auf.
 - F: Hier ist eine deutlich fehlerhafte Tiefenzuordnung zu erkennen. Ursache dafür ist vermutlich der helle Bereich, welcher durch die Beleuchtung verursacht wird. Dieser Bereich wird als Szenenelement interpretiert, daher wird ihm eine falsche Tiefe zugeordnet.
 - G: Die Struktur des Statives wird gut erkannt, und auch die einzelnen Tiefenwerte werden relativ verlässlich zugeordnet.
 - H: Der Verlauf der Pappröhre wird gut erfasst, dies gilt für Gestalt ebenso wie für die Tiefe.
- Wand 03_3

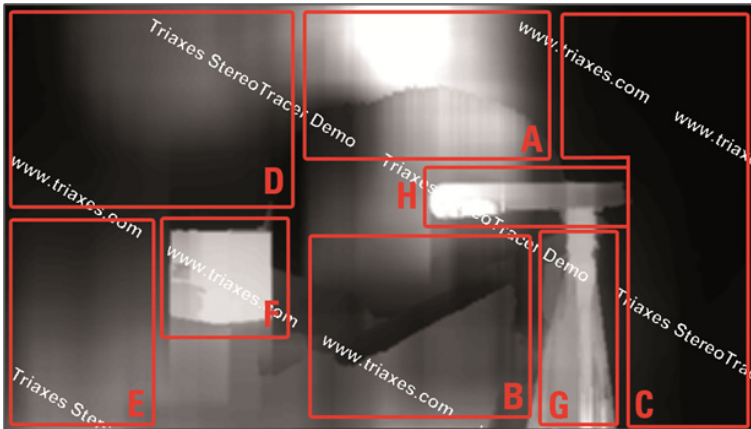


Abbildung 35: TK Analyse Wand 03_3

- A: Es werden unterschiedliche Tiefenebenen für ein und dieselbe Fläche ermittelt. Mögliche Erklärung: Auch hier bewirkt die Beleuchtung eine unterschiedliche Helligkeitsverteilung, was zu einer falschen Zuordnung der Tiefenwerte führt.
- B: Hier führt der Schattenwurf der Pappröhre zu einer falschen Zuordnung von Tiefenwerten.
- C: Für den Hintergrund ist die Tiefe korrekt zugeordnet.
- D: Der Tiefenverlauf in diesem Bereich entspricht nicht der tatsächlich vorhandenen Tiefe. Mögliche Erklärung: Die Beleuchtung verursacht Helligkeitsunterschiede, welche falsch interpretiert werden.
- E: Der Bereich weist insgesamt recht ähnliche Tiefenwerte auf, diese sind jedoch im Vergleich zur Tiefe in Bereich C nicht passend, da sie eigentlich identisch sein müssten.
- F: Die Struktur der Schachtel wird gut erfasst, allerdings gilt dies nur bedingt für den Tisch, bzw. die Tischplatte. Auch der Tiefenwert für die Schachtel wird relativ genau erfasst, obwohl keine Trennung zwischen den beiden verschiedenen Flächen erfolgt. Mögliche Erklärung: Zum Hintergrund erfolgen markante Hell-Dunkel-Übergänge. Beim Übergang zur Tischplatte sind diese Übergänge wesentlich weniger intensiv.
- G: Der Tiefenverlauf der Pappröhre wird relativ gut erfasst. Mögliche Erklärung: Durch die Beleuchtung erfolgt eine scharfe Abgrenzung des Objekts zum Hintergrund.

4 Zusammenfassung und Fazit

Schon bei der Aufnahmen erstellte Tiefenkarten können in der S3D-Postpro wie auch bei der Distribution einen segensreiche Wirkung entfalten und diesen Prozessschritten einen deutlichen (kreativen und ökonomischen) Effizienzscharb verleihen.

In diesem Artikel wurde gezeigt, welche Grundlagen für die Gewinnung von Tiefeninformationen aus Stereobildern relevant sind. Es wurden die damit zusammenhängenden mathematischen Beziehungen erläutert, sowie die in diesem Prozess involvierten technischen Komponenten vorgestellt. Desweiteren wurden konkrete Verfahren zur Gewinnung von Rauntiefe erklärt, welche auf diesen Rahmenbedingungen beruhen.

Den geringsten Hardwareaufwand zur Generierung von Tiefenkarten bietet in der S3D-Produktion die Stereo-Triangulation. Diese wurde daher nochmals genauer betrachtet. Für die Stereotriangulation kann auf eine große Anzahl bereits existierender Programmibliotheken zurückgegriffen werden, ein Beispiel hierfür ist das Integrating Vision Toolkit (kurz: IVT)¹⁰⁰, welches am Karlsruher Institut für Technologie (kurz: KIT) entwickelt wurde. Zur Generierung unserer Tiefenkarten wurde das Programm Stereotracer verwendet, das die Verarbeitung einzelner Stereobildpaare oder Sequenzen von Einzelbildpaaren erlaubt.

Auf der Grundlage von Testaufnahmen und deren Resultate konnten die Potenziale und Grenzen für einen praktischen Systemaufbau erarbeitet werden. Für die Testaufnahmen wurden zwei Panasonic-Camcorder NV-GS500 in einer achsparallelen Ausrichtung verwendet. Die Aufnahme erfolgte im Standard-PAL-Format bei einem Seitenverhältnis von 16:9. Über die Aufnahmen sollten möglichst viele unterschiedliche Szenensituationen abgedeckt und bewertet werden können. Die jeweiligen Szenensequenzen wurden vorab mit den Adobe-Programmen Premiere und After Effects weitgehend von vertikalen Disparitäten befreit.

Die theoretischen Grenzbetrachtungen zur Stereotriangulation wurden durch die Tests weitestgehend bestätigt. Deutliche Objektkanten können hilfreich sein Tiefenebenen zu identifizieren. Für kontrastarme Bereiche wird oft eine falsche Zuordnung ermittelt. Stärker als erwartet scheint sich die Beleuchtung auszuwirken. Der Genauigkeit einer Tiefenkarte dürfte daher eine professionelle Szenenbeleuchtung, mit der die Tiefe betont wird sehr zuträglich zu sein. Verdeckungen und Tiefensprünge erweisen sich als einflussreiche Fehlerquellen. Insgesamt werden die Strukturen und Tiefenebenen zu grob erfasst, Details gehen sehr schnell verloren. Zum Teil treten für unterschiedliche Bildinhalte erhebliche Diskrepanzen bei der Zuordnung der korrekten Tiefenwerte auf, so dass für professionelle Anwendungen in der S3D-Filmproduktion evtl. doch ergänzende Verfahren jenseits der visuellen Stereotriangulation notwendig werden.

100 Vgl. Azad 2010.

Literaturverzeichnis

Monografien

- Bender, Michael; Brill, Manfred (2006): Computergrafik. Ein anwendungsorientiertes Lehrbuch. 2., überarb. Aufl. München: Hanser.
- Dobbert, Tim (2005): Matchmoving. The invisible art of camera tracking. San Francisco, Calif., London: SYBEX.
- Jähne, Bernd (2005): Digitale Bildverarbeitung. 6., überarbeitete und erweiterte Auflage. Berlin, Heidelberg: Springer.
- Mach, Rüdiger; Petschek, Peter (2006): Visualisierung digitaler Gelände- und Landschaftsdaten. Berlin, Heidelberg: Springer.
- Meschede, Dieter (2008): Optik, Licht und Laser. 3., durchges. Aufl. Wiesbaden: Vieweg + Teubner.
- Neumann, Burkhard (2005): Bildverarbeitung für Einsteiger. Programmbeispiele mit Mathcad. Berlin Heidelberg: Springer
- Nischwitz, Alfred; Fischer, Max; Haberäcker, Peter (2007): Computergrafik und Bildverarbeitung. Alles für Studium und Praxis – Bildverarbeitungswerkzeuge. 2., verb. u. erw. Aufl. Wiesbaden: Vieweg.
- Schreer, Oliver (2005): Stereoanalyse und Bildsynthese. Berlin, Heidelberg: Springer.
- Steinmüller, Johannes (2008): Bildanalyse. Von der Bildverarbeitung zur räumlichen Interpretation von Bildern. Berlin /// Berlin, Heidelberg: Springer (eXamen).
- Wöhler, Christian (2009): 3D Computer Vision. Efficient Methods and Applications. Berlin: Springer (X).

Fachartikel

- Bobick, Aaron F.; Stephen S. Intille (1999): Large Occlusion Stereo. In: International Journal of Computer Vision, Jg. 33, H. 3, S. 181–200.
- Canny, John (1986): A Computational Approach to Edge Detection. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, H. 8, S. 679–698.
- Faugeras, Olivier et al. (1993): Real time correlation-based stereo: algorithm, implementations and applications. Unter Mitarbeit von Bernard Hotz, Herve Mathieu, Thierry Vieville et al. Institut National de Recherche en Informatique et en Automatique (INRIA).
- Gu, C.; Wu, L. (1990): Structural Matching of Multiresolution for Stereo Vision. International Conference on Pattern Recognition. Unter Mitarbeit von.
- Hamming, R. W. (1950): Error Detecting and Error Correction Codes. In: The Bell System Technical Journal, Jg. 29, H. 2.
- Harris, Chris; Stephens, Mike (1988): A Combined Corner and Edge Detector. Plessey Research Roke Manor, United Kingdom.
- Hile, Harlan; Zheng Colin (2004): Stereo Video Processing for Depth Map. Uni-

versity of Washington.

- Hourad, R.; Skordas, T. (1998): Structural Matching for Stereo Vision. 9th International Conference on Pattern Recognition.
- Kanade, Takeo; Okutomi, Masatoshi (1991): A Stereo Matching Algorithm with an Adaptive Window. Theory and Experiment. Proceedings of the IEEE International Conference on Robotics and Automation.
- Kawakita, M. et al. (2002): HDTV Axi-Vision Camera. NHK Science & Technical Research Laboratories (Japan). Japan.
- Kawakita, Masahiro et al. (2004): High-definition real-time depth-mapping TV camera: HDTV Axi-Vision Camera. Optical Society of America.
- Klaus, Andreas; Sormann, Mario; Karner, Konrad (2006): Segment-Based Stereo Matching Using Belief Propagation and a Self-Adapting Dissimilarity Measure. VRVis Research Center, 8010 Graz, Austria.
- Moravec, Hans P. (1981): Rover Visual Obstacle Avoidance. Proceedings of the seventh International Joint Conference on Artificial Intelligence.
- Niederoest, Markus; Niederoest, Jana; Šcucka, Jirí (2003): Shape from Focus: Fully Automated 3D Reconstruction and Visualization of Microscopic Objects. Institute of Geodesy and Photogrammetry, ETH Zurich, Switzerland; Institute of Geonics, Academy of Sciences of the Czech Republic, Ostrava, Czech Republic.
- Nillius, Peter; Eklundh, Jan-Olof (2002): Fast Block Matching with Normalized Cross-Correlation using Walsh Transforms. Department of Numerical Analysis and Computer Science, Royal Institute of Technology Stockholm, Sweden.
- Psarakis, Emmanouil Z.; Evangelidis, Georgios D. (2005): An Enhanced Correlation-Based Method for Stereo Correspondence with Sub-Pixel Accuracy. In: Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05).
- Schreer, Oliver; Hartmann, Irmfried; Adams, Roger (1997): Analysis of Grey-Level Features for Line Segment Stereo Matching. Institute for Measurement and Automation Technology, Department of Electrical Engineering, Technical University Berlin. In: Lecture Notes in Computer Science, S. 620–627.
- Smolic, Aljoscha et al. (2008): Intermediate view interpolation based on multi-view video plus depth for advanced 3D video systems. IEEE International Conference on Image Processing, ICIP 2008. Conference Proceedings. Piscataway NJ: IEEE.
- Vincze, M. et al. (2008): Roboternavigation in Büros und Wohnungen. In: Elektrotechnik und Informationstechnik, Vol. 125, No. 1-2 (2008), S. 25–32.
- Zabih, Ramin; Woodfill, John (1994): Non-parametric Local Transforms for Computing Visual Correspondence. Computer Science Department, Cornell University, Ithaca NY; Interval Research Corporation, Palo Alto CA.
- Zhan, Pengcheng; Lee, Dah-Jye; Beard, Randal (2004): Solving correspondence problem with 1D signal matching. Brigham Young University, Department of Electrical and Computer Engineering.
- Zilly, Frederik; Eisert, Peter; Kauff, Peter (2009): Real-Time Analysis and Cor-

rection of Stereoscopic HDTV Sequences. Fraunhofer-Institut für Nachrichtentechnik, Heinrich-Hertz-Institut (Berlin).

- Hochschulschriften
- Ebers, Olga (2004): Überblick über aktuelle Verfahren zur Tiefenschätzung aus 2D-Video-Sequenzen. Studienarbeit. Berlin. TU Berlin, Institut für Telekommunikationssysteme, FG Nachrichtenübertragung.
- Frank, Anja (2007): Stereoskopische High-Definition Produktion. Bachelor-Thesis. Betreut von Nikolaus Hottong und Daniel Fetzner. Furtwangen. Hochschule Furtwangen University HFU, Fakultät Digitale Medien.
- Internetquellen
- 2D-plus-depth. Online verfügbar unter <http://en.wikipedia.org/wiki/2D-plus-depth>, zuletzt geprüft am 25.05.2010.
- Azad, Pedram (2010): Integrating Vision Toolkit (IVT). Karlsruhe Institute of Technology (KIT). Online verfügbar unter: <http://ivt.sourceforge.net/>, zuletzt geprüft am 28.05.2010
- RP Photonics Consulting GmbH (Hg.): White Light Interferometers. Online verfügbar unter http://www.rp-photonics.com/white_light_interferometers.html, zuletzt geprüft am 25.05.2010.

Sonstige Veröffentlichungen

- Hallpike, Lucy (2009): The Role of Ocula in Stereo Post Production. The Foundry Visionmongers Ltd.
- SMPTE (Hg.): Society of Motion Picture and Television Engineerings Inc. (2009): Report of SMPTE Task Force on 3D to the Home. Society of Motion Picture and Television Engineerings Inc. (SMPTE).
- Scharfenberg, Harald (2009): Stereo Vision. Korrespondenzanalyse. Skript zum Masterstudiengang Optotechnik und Bildverarbeitung. Hochschule Darmstadt.
- Smirnov, Sergey et al. (2010): 3D Video Processing Algorithms - Part 1. Technical report D5.4. Herausgegeben von Mobile 3DTV.
- Waldowski, Michael (Hg.) (2010): Skript Grafische Datenverarbeitung. Unter Mitarbeit von André Bohnet-Armstrong, Franziska Barth und Kathrin Kühnle et al. Hochschule Furtwangen University.

Teil 6

Untersuchungen zur H264 Multi-view Video Coding Extension für den Einsatz in der Distribution von S3D Filmen

Martin Maurer

Abbildungsverzeichnis

Abbildung 1: Aufbau eines DVD Breitbilds (Europa).....	351
Abbildung 2: Aufbau eines monoskopischen Bluray-Disc Bildes.....	352
Abbildung 3: Side-By-Side-Verfahren.....	353
Abbildung 2.6.: Side-By-Side-Squeezed Verfahren	354
Abbildung 2.7.: Over-and-Under-Verfahren	355
Abbildung 2.8.: Over-and-Under-Squeezed-Verfahren	356
Abbildung 2.9.: Field-Sequential.....	356
Abbildung 2.10.: Simulcast.....	357
Abbildung 2.11.: Video plus Depthmap	357
Abbildung 4: HDMI 1.4- Frame Packing Format.....	358
Abbildung 3.1.: Free Viewpoint Video.....	360
Abbildung 3.2.: MVC-Architektur	362
Abbildung 4.1.: H264/AVC-Encoder	367
Abbildung 4.2.: Aufbau einer H264/AVC NAL-Unit.....	367
Abbildung 4.3.: Interview Prediktion	369
Abbildung 4.4.: Aufbau NAL-Unit eines Baseview	370
Abbildung 5.1.: Aufruf des MVC-Encoders für View 0 und View 1.....	373
Abbildung 5.2.: Aufruf des MVCBitStreamAssemblers	376
Abbildung 5.3.: Aufruf des H264AVCDecoderLibTestStatic	377
Abbildung 5.4.: Aufruf des PSNRStatic Tools	377
Abbildung 5.5.: Samples Microsoft Research 3D	378
Abbildung 5.6.: Samples MERL	379
Abbildung 6.1.: Visueller Vergleich QP 5 bis QP 20.....	391
Abbildung 6.2.: Visueller Vergleich QP 20 bis QP 40	392
Abbildung 6.3.: Visueller Vergleich QP 20 bis QP 40	396

Tabellenverzeichnis

Tabelle 5.1.: Zugangsdaten für das JMVC CVS-Repository der RWTH Aachen	372
Tabelle 5.2.: Übersicht über wichtigsten Parameter der Konfigurationsdatei	374
Tabelle 6.1.: Vergleich Y-PSNR (Chrominanz)- Ballettszene	385
Tabelle 6.2.: Vergleich U-PSNR (Luminanz U) - Ballettszene	385
Tabelle 6.3.: Vergleich V-PSNR(Luminanz V) - Ballettszene	385
Tabelle 6.4.: Vergleich Y-PSNR (Chrominanz)- Ballroomsequenz	386
Tabelle 6.5.: Vergleich U-PSNR (Luminanz U) - Ballroomsequenz	386
Tabelle 6.7.: Vergleich Simulkast und Interview Prediktion für GOP=8, GOP=16 und GOP=32	387
Tabelle 6.8.: Vergleich GOP=8, GOP=16 und GOP=32	388

1 Einführung

Der stereoskope Film erlebt im Kino zurzeit besonders starke Beachtung. Leider hat sich dieser Trend aktuell noch nicht ernsthaft bis in den Heimbereich fortgesetzt, da u.a. viele Interoperationalitäts- und Kompatibilitätsfragen derzeit noch unbeantwortet sind.

Mit der Verabschiedung von H.264/MVC als offizieller Bluray-Disc-Codec, ist erstmalig ein herstellerübergreifender Codieransatz für stereoskopische HD-Videomaterial verfügbar, der aufgrund der monoskopischen Abwärtskompatibilität zu H264/AVC eine große Chance hat, sich am Markt zu etablieren.

Dieser Artikel liefert einen ersten Qualitätsüberblick über die Multi-view Video Coding Extension (MVC) des H.264/AVC-Videostandards und ihre Anwendung im Bereich der 3D-Bluray und des Free View Point Television.

Es wird im ersten Schritt zunächst ein Überblick über die Wiedergabe- und Speicherprinzipien für stereoskope Bildinhalte vorgestellt.

Im zweiten Schritt wird der Bereich des Multi-view Videos dargestellt und darauf basierend die beiden Anwendungsbereiche 3D-TV und Free View Point Video (FVP) erläutert. Zusätzlich werden die Architektur und die Anforderungen an Multi-view Video aufgezeigt. Anschließend werden die Veränderungen der Multi-view Video Coding Erweiterung gegenüber dem H.264/AVC-Videostandard dargelegt und der JMVC-Referenzcodec vorgestellt.

In einem einfachen Testdurchlauf wurde dieser Codec mit mehreren, frei-verfügbaren S3D-Videosequenzen getestet und ein erster Bezug zwischen Zieldatenrate, Signal-

Rauschabstand (PSNR) des Codierergebnisses und subjektiver Bildqualität hergestellt. Ziel war es, eine erste grobe Einschätzung des Codecs bez. seiner Fähigkeiten zur Distribution von S3D-Material zu bekommen.

Es sei darauf hingewiesen, dass es sich hierbei nicht um standardisierte Qualitätsbetrachtungen in korrekter empirischer Methodik handelt, sondern um eine erste subjektive Annäherung an den JMVC-Referenzcodec zur Vorbereitung weiterer Tests.

2. Stereoskope Speicher- und Übertragungsprinzipien

Zur Speicherung der stereoskopischen Bildinformationen auf Bluray-Disc und DVD haben sich mehrere Verfahren etabliert. Jedes Verfahren hat seine Vorteile und Nachteile in Bezug auf Darstellungsqualität, Speichereffizienz und Abwärtskompatibilität.

Auch sind die Speicherkapazitäten von Bluray-Discs und Video-DVDs begrenzt. Eine doppelagige PAL Video-DVD fasst 8,5GB an Videodaten, welche in MPEG2 mit einer maximalen Datenrate von 9,8Mbit/s und 720 Pixel horizontal und 576 Pixel vertikal gespeichert sind¹. Videoinhalte in Widescreen können sowohl letterboxed als auch anamorph codiert werden. Der DVD-Video-Standard sieht keine offizielle Erweiterung für stereoskopische Bildinhalte vor, so dass neben dem optionalen Videostandard Sensio zahlreiche DVDs mit unterschiedlichen Verfahren auf dem Markt sind.

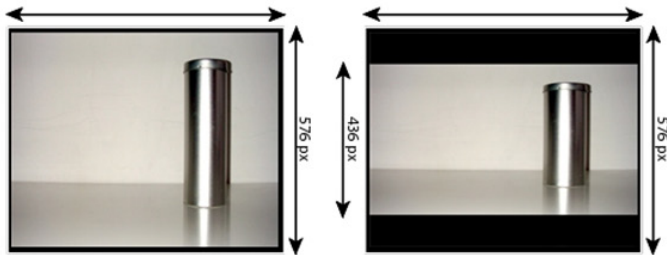


Abbildung 1: Aufbau eines DVD Breitbilds (Europa)

Eine doppelagige Bluray-Disc fasst 50GB an Videodaten, während eine einlagige Bluray Disc 25GB fasst. Eine Bluray Disc kann sowohl Videoinhalte in MPEG2, VC1 als auch H264/AVC Codierung beinhalten, deren maximale Videobitrate auf 40 Mbit/s begrenzt ist. Die maximale spatiale Auflösung einer Bluray-Disc ist auf 1920 Pixel horizontal und 1080 Pixel vertikal bei einer maximalen Bildwiederholrate von 60 Hertz begrenzt.

Ende des Jahres 2009 wurde die Multiview-Video-Coding Extension zur Speicherung von stereoskopischen Inhalten in den Bluraystandard aufgenommen². MVC-codierte Bluray-Discs abwärtskompatibel sein und monoskopisch auf älteren Bluray-Standaloneplayern abspielbar sein.

¹ Vgl. Taylor, DVD Demystified, (2001), S. 09-35.

² Vgl. Blu-ray Disc Association, Blu-ray Disc Association Provides Updates on 3D Technology, (2009).

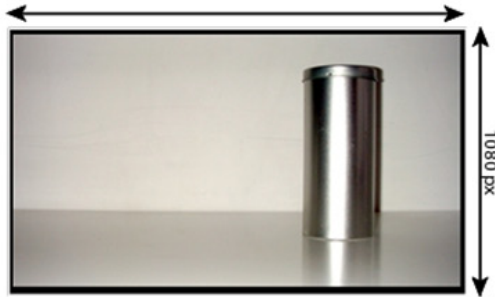


Abbildung 2: Aufbau eines monoskopischen Blu-ray-Disc Bildes

Bevor MVC als offizielle Blu-ray-3D-Erweiterung verabschiedet wurde, waren schon über 30 stereoskopische Filme auf Blu-ray-Disc erschienen. Sie waren durchweg anaglyphisch in 1920x1080 px codiert.

2.1. DCDM

DCDM steht für Stereoscopic Digital Cinema Abbendum und ist eine Erweiterung des bisherigen DCI-Standards (Digital Cinema Initiative) zur digitalen Speicherung von stereoskopischen Videomaterial im Kino. Die Spezifikationen wurden am 11. Juli 2007 verabschiedet.

DCDM spezifiziert, dass nur ein digitales Filmmaster, das sogenannte Distributions Master (DCDM) pro stereoskopischen Film erstellt wird und dieses keine Optimierungen für ein bestimmtes stereoskopisches Wiedergabematerial enthält. Weiterhin soll nur ein Distribution Package (DCP) pro stereoskopischen Film angefertigt werden, welches einen Videostream mit der Bandbreite von 250 Mbit/s enthält. In diesem werden der linke und rechte Blickwinkel, alternierend bei 48 Frames pro Sekunde, gespeichert, mit dem linken Blickwinkel zuerst. Außerdem soll nur ein Audiostream für monoskopische und stereoskopische Wiedergabe verwendet werden³.

2.2. Side-by-Side

Beim Side-by-Side-Verfahren werden beide stereoskopischen Ansichten nebeneinander in einem einzelnen Videobild gespeichert und übertragen. Einerseits gibt es das normale Side-by-Side-Verfahren, bei welchem beide Videobilder in voller Auflösung gespeichert werden. Dadurch verdoppelt sich die benötigte horizontale Auflösung des zusammengesetzten Videobildes auf 3840 Pixel bei FullHD-Auflösung und 1440 Pixel bei PAL-Auflösung.

³ Vgl. Mendiburu, 3D Movie Making, (2009), S.187f.

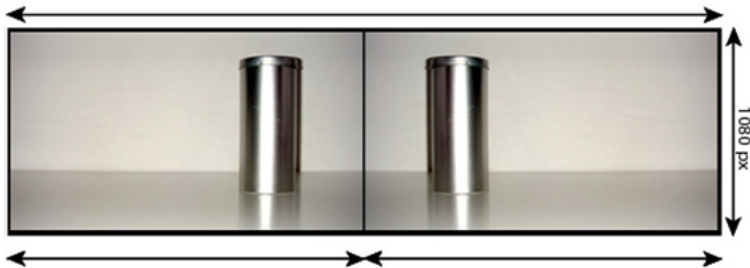


Abbildung 3: Side-By-Side-Verfahren

Andererseits gibt es auch das Side-By-Side-Squeezed-Verfahren, welches beide Ansichten gestaucht im Videoframe speichert⁴⁵. Dadurch verringert sich allerdings die horizontale Auflösung auf die Hälfte, während die vertikale Auflösung gleich bleibt. Auf einer Video-DVD stehen dann nur noch 360 horizontale Bildpunkte zur Verfügung und auf Bluray-Discs nur 960 Pixel.

Das Side-By-Side-Squeezed-Verfahren ist das einzige side-by-side 3D-Verfahren, dass gemäß der Bluray-Spezifikation in traditioneller Codierung möglich ist, da die Größe des Videobildes auf der Bluray-Disc auf maximal 1920 Pixel horizontal zu 1080 Pixel vertikal begrenzt ist. Das Verfahren hat natürlich einige Nachteile: Neben dem horizontalen Auflösungsverlust (gerade bei Bildmaterial, welches vornehmlich auf Querdisparitäten abzielt) ist auch die Kompatibilität zu herkömmlichen monoskopischen Standalone-Playern nicht gegeben. Diese würden nur ein gestauchtes Bild mit linkem und rechtem Halbbild gleichzeitig darstellen, was für den Zuschauer ein eher unbefriedigendes Seherlebnis darstellt. Nur am PC mit geeigneter Software, wie dem Stereoscopic Player ist die Wiedergabe problemlos möglich.

2.3. Over-and-Under

Beim Over-and-Under-Verfahren werden beide Ansichten übereinander in einem einzelnen Videoframe gespeichert und übertragen. Einerseits gibt es das normale Over-and-under Verfahren, bei welchem beide Ansichten in voller Auflösung gespeichert werden. Dadurch verdoppelt sich die benötigte vertikale Auflösung des zusammengesetzten Videobilds auf 2160 Pixel vertikal bei FullHD-Auflösung und 1152 Pixel bei PAL-Auflösung.

Andererseits gibt es auch das Over-and-Under-Squeezed-Verfahren, welches beide An-

4 Vgl. Janssen, c't 15/2009 - Räumlich am Rechner, (2009b).

5 dto

sichten gestaucht in einen Videoframe presst. Dadurch verringert sich allerdings die vertikale Auflösung auf die Hälfte, während die horizontale Auflösung gleich bleibt. Auf einer Video-DVD stehen dann nur noch 288 vertikale Bildpunkte zur Verfügung und auf Bluray-Discs nur 540 Pixel. Das Over-and-Under-Squeezed-Verfahren ist das einzige Over-and-Under-Verfahren, dass gemäß der Bluray-Spezifikationen in traditioneller Codierung möglich ist, da die Größe des Videobildes auf der Bluray-Disc auf maximal 1920 Pixel horizontal zu 1080 Pixel vertikal begrenzt ist.

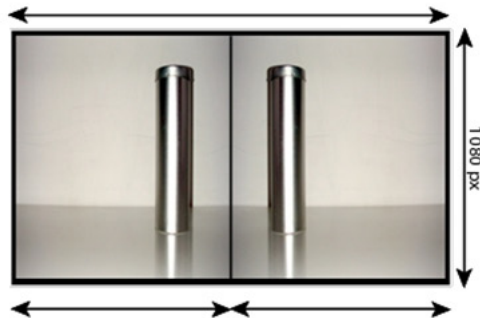


Abbildung 2.6.: Side-By-Side-Squeezed Verfahren

2.4. Field-Sequential/Interlaced-Stereo

Das Field-Sequential-Verfahren ist eines der ältesten Speicher- und Übertragungsverfahren für stereoskopisches Video. Bei dieser Technik werden beide Blickwinkel in Zeilensprung-basierten Halbbildern, den sogenannte Fields, gespeichert. Diese Vorgehensweise ist nur bei nativ im interlace-Format vorliegendem Video sinnvoll.

Durch die Speicherung in beiden Halbbildern eines Videoframes halbiert sich die vertikale Auflösung, sodass auf DVD nur 288 Pixel pro Halbbild bzw. Blickwinkel zur Verfügung stehen. Weiterhin ist die Bildwiederholrate von herkömmlichen PAL und NTSC-Geräten mit 50 Hertz bzw. 60 Hertz für die Betrachtung mit Shutterbrillen zu gering⁶.

⁶ Vgl. Janssen, c't 15/2009 - Räumlich am Rechner, (2009b).

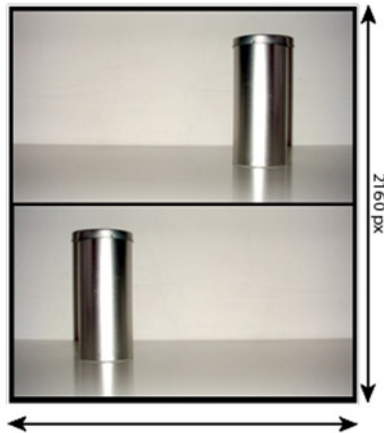


Abbildung 2.7.: Over-and-Under-Verfahren

2.5. Sensio

Sensio ist eine kanadische Firma die Video-DVDs und Video-Decoder anbietet. Sensio 3D wurde im Januar 2009 offiziell als optionale Erweiterung für stereoskopische Video-DVD in den DVD-Standard aufgenommen⁷. Es gibt zwei Speicherverfahren, die von Sensio vermarktet werden. Unter der Bezeichnung Sensio HQFS werden normale Field-interlaced codierte Filme angeboten, während unter der Bezeichnung Sensio 3D Video-DVDs vertrieben werden, die ein proprietäres Side-By-Side Verfahren benutzen⁸.

Bei diesem Verfahren werden vor der Speicherung in Form eines Schachbrettmusters die Hälfte der Pixel eines Videobildes entfernt. Das bedeutet, dass in der ersten Zeile alle geraden Bildpunkte verworfen werden und in der zweiten Zeile alle ungeraden Bildpunkte. Dies erfolgt für das linke und rechte Videobild. Danach werden beide Videobilder nebeneinander im Videoframe gespeichert⁹. Bei der Decodierung werden die fehlenden Bildpunkte durch den Sensio Decoder künstlich hinzugerechnet. Laut Herstellerangaben soll so nur 30% statt 50% der Auflösung verloren gehen¹⁰. Die Zukunft dieses Verfahrens ist ungewiss. Einerseits sind seit der Implementierung als offizielle Erweiterung für DVD-Video vor einem Jahr keine neuen Filme in diesem Format erschienen. Andererseits ist das bisherige Filmangebot auf ältere Filmklassiker und eine geringe Anzahl an Erotikfilmen beschränkt¹¹.

7 Vgl. Janssen, 3D-Standard für DVD, (2009a).

8 Vgl. <http://www.stereo3d.com>, <http://www.stereo3d.com> 2005 - Sensio format, (2005).

9 Vgl. www.stereoforum.org, Anaglyphes 3D in zwei Streams konvertieren, (2010).

10 Vgl. Janssen, c't 15/2009 - Räumlich am Rechner, (2009b).

11 siehe http://www.sensio.tv/en/home_theater/3d_dvd/available/default.3d

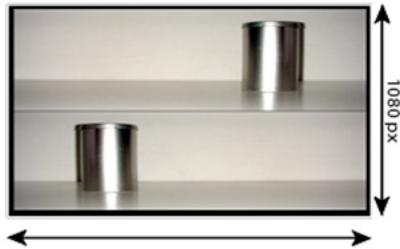


Abbildung 2.8.: Over-and-Under-Squeezed-Verfahren

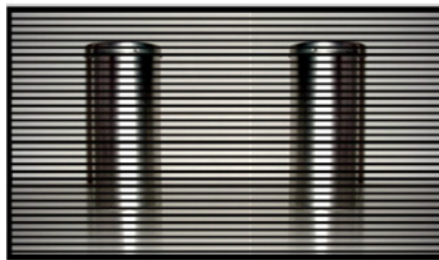


Abbildung 2.9.: Field-Sequential

2.4.7. Simulcast

Beim Simulcast-Verfahren werden das rechte und linke Videobild jeweils als separater Videostrom in einem Videocontainer (z.B. MPEG Transportstream) gespeichert. Dadurch bleibt die vollständige Bildauflösung und Framerate beider Ansichten erhalten. Auch verändert sich die Güte der Farbwiedergabe nicht.

Der Nachteil des Simulcastverfahren ist aber der gesteigerte Ressourcenbedarf bei der Übertragung und Wiedergabe durch die verringerte Codiereffizienz. Während es bereits auf der DVD in Form von Multi Angle die Möglichkeit gab, mehrere Videostreams parallel zu speichern und zwischen diesen nacheinander umzuschalten, gibt es bisher keine Referenzimplementierung, bei welcher diese Videostreams parallel für stereoskopisches Video decodiert werden können¹².

¹² Vgl. Mendiburu, 3D Movie Making, (2009), S. 193

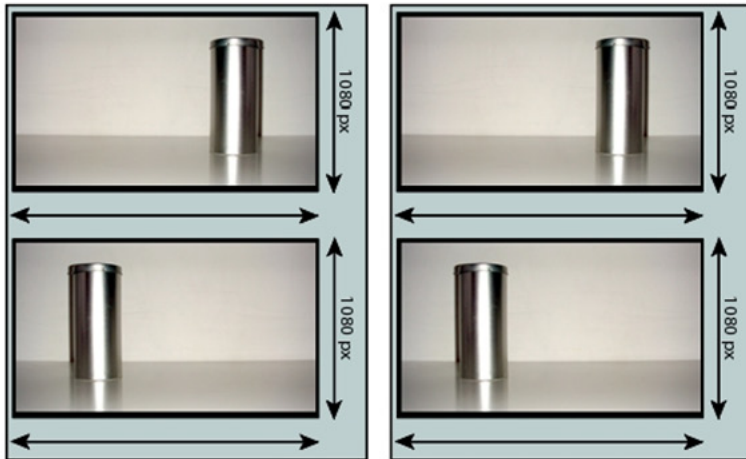


Abbildung 2.10.: Simulcast

2.4.8. Video plus Depthmap



Abbildung 2.11.: Video plus Depthmap

Beim Video plus Depthmap-Verfahren (manchmal auch als 2D plus Deph bezeichnet) wird neben dem Videobild zusätzlich eine Depthmap codiert und gespeichert. Diese Depthmap enthält Graustufenwerte und kann so sehr effektiv codiert werden (nähere Details dazu sind in diesem Arbeitspapier im Artikel von Eric Schumacher zu finden). In MPEG2 gibt es eine Referenzimplementierung für Video plus Depthmap. Das reguläre Videobild wird als Base Layer codiert und kann so von herkömmlichen MPEG2-Decodern decodiert werden.

Zusätzlich wird ein zweiter Layer eingefügt. Dieser Depth Layer enthält die Depthmap, welche nur von kompatiblen Decodern decodiert werden kann¹³.

¹³ Vgl. Ozaktas/Onural, Three-Dimensional Television, (2008), S. 325f.

2.6 Schnittstellt HDMI

Das High Definition Multimedia Interface ist zur Übertragung digitaler Audio-/Video-Signale zwischen verschiedenen Consumer-Elektronikgeräten konzipiert. Die sieben bei der Entwicklung beteiligten Firmen waren Hitachi, Matsushita/Panasonic, Philips, Silicon Image, Sony, Thomson/Technicolor und Toshiba. HDMI ist bidirektional ausgelegt. Einer der wichtigsten Gründe hierfür ist die wechselseitige Übertragung von Steuersignalen zwischen den verbundenen Heimgeräten. Die Übertragung von Audiosignalen kann auf bis zu acht Kanälen erfolgen. HDMI unterstützte dabei von Anbeginn an wichtige Audiokonzepte wie etwa sämtliche Dolby-Formate (Dolby Digital / Dolby 5.1 audio etc.) sowie DTS-Formate (Digital Theater Systems Mehrkanaltonsystem) mit komprimierten Audiodaten¹⁴. Vereinzelter Kritikpunkt ist allerdings die zulässige maximale elektrische Kabellänge von 15m.

Mit Version HDMI 1.3 wurde 2006 die verfügbare Videobandbreite auf 10,2 GBit/sec mehr als verdoppelt und die maximal mögliche Farbtiefe pro Pixel wurde auf bis zu 45bit heraufgesetzt.

Mit der Version 1.4 von 2010 wurde ein 4k-Einzelbild-Framing bis zu 3840x2160 pixel definiert, was folglich auch die gleichzeitige Übertragung von zwei vollen HD-Ansichten in einem solchen Megaframe erlaubt. Konsequenterweise wurde daher die Spezifikation auf mehrere Stereo-HD-Übertragungsformate erweitert u.a. auf das hochauflösende, progressive Frame-Packing-Format und diverse 2D plus Depth HD-Varianten¹⁵.

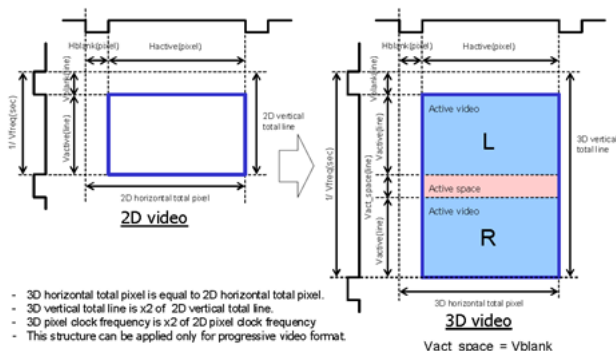


Abbildung 2.21: HDMI 1.4- Frame Packing Format

¹⁴ <http://www.hdmi.org/>

¹⁵ High-Definition Multimedia Interface Specification Version 1.4a - Extraction of 3D Signaling Portion

3. Multi-view Video

3.1. Definition

Unter dem Begriff Multi-view Video versteht man eine n-Anzahl von zeitlich synchronisierten Videoströmen, die das gleiche Objekt von einer n-Anzahl von Ansichten (sogenannten Views) aufnehmen und darstellen. Besondere Anwendung findet Multi-view Video im Bereich 3D-Video (3DV) und Free Viewpoint Video (FVV)¹⁶.

3.2. Historische Entwicklung

Erste Überlegungen der Moving Picture Experts Group (MPEG) zur Codierung dreidimensionaler Videosignale begannen im Dezember 2001 mit Gründung der 3D Audio-Visual (3DAV)-Arbeitsgruppe durch die MPEG¹⁷. Diese sammelte zwischen Oktober 2003 und Dezember 2005 Vorschläge für die Standardisierung der Multi-view Video Erweiterung und evaluierte diese Vorschläge auf dem 75. MPEG-Treffen 2006 in Bangkok¹⁸. Die Ergebnisse flossen in die Ausarbeitung der Anforderung an den Standard und den Referenzcodec JMVC ein. Im Juli 2006 wurde auf dem 77. MPEG-Treffen in Klagenfurt beschlossen, dass die Standardisierung durch das Joint Video Team (JVT) durchgeführt werden soll¹⁹. Das JVT ist eine Vereinigung von Videoexperten der ITU-T VCEG (Video Coding Experts Group) und der Teile der MPEG.

3.3. Anwendungsszenarien

3.3.1. Free Viewpoint Video

Der Begriff Free Viewpoint Video bezeichnet ein Anwendungsszenario in welchem der Benutzer seinen Blickwinkel und Standpunkt in der Bildszene frei bestimmen kann. Dies ermöglicht einen hohen Grad an Interaktivität, bedeutet aber auch eine gesteigerte Anforderung an das Aufnahme- und Verarbeitungssystem. Am Besten eignet sich das Verfahren für Vorgänge, die sie sich in einem vordefinierten und begrenzten Bereich abspielen, z.B. Stadion bei einer Sportveranstaltung. Wenn ein Stadion im Vorfeld komplett mit Kameras um das Spielfeld ausgestattet wird, kann sich der Zuschauer virtuell während der Veranstaltung kann auf dem Spielfeld bewegen²⁰.

Einfache Implementierungen dieses Verfahrens sind bereits im Einsatz. z.B. Formel 1-Übertragung, beschränken sich aber auf einfache einfaches Umschalten zwischen einzelnen Kamerabildern²¹. Auch filmisch wurde dieses Verfahren im Jahr 1999 mit dem Film The Matrix als Bullet Time-Effekt weltbekannt. Für diesen Effekt wurden

16 Vgl. Ozaktas/Onural, Three-Dimensional Television, (2008), S. 314

17 Vgl. Martinez-Ibañez, MPEG 3DAV AhG Activities Report, (2003), S. 1.

18 Vgl. Pereira et al., Distributed Coding for Video Services, (2006), S. 5.

19 dito

20 Vgl. Eisert et al., Free View Point Video, (2006), S. 1.

21 Vgl. Ozaktas/Onural, Three-Dimensional Television, (2008), S. 319f.

kreisförmig um den Schauspieler zahlreiche Kameras angebracht, die den Schauspieler synchronisiert filmten.

In der Postproduction konnte später beliebig zwischen den Kameras gewechselt werden und so eine virtuelle Kamerafahrt erstellt werden. Nachteil dieser Methode ist allerdings, dass die virtuelle Kamerafahrt vor der Produktion sehr genau geplant werden muss, da an dieser die einzelnen Kameras ausgerichtet werden müssen. Wenn später Änderungen an der virtuellen Kamerafahrt vorgenommen werden müssen, ist dies meist nur durch Neuaufnahme der kompletten Szene möglich²².

Generell ist es für eine glaubwürdige Szenenwiedergabe wünschenswert so viele reale Kamerabilder wie möglich zu benutzen. Eine beliebige Steigerung der Anzahl der Kameras ist aus wirtschaftlichen und auch physikalischen Gründen nicht möglich²³. Auch steigt der Synchronisationsaufwand mit der Anzahl der Kameras. Daher werden verschiedene Verfahren benutzt, um eine glaubwürdige Wiedergabe der Szene aus einer beschränkten Anzahl an physischen Kameras und Views zu ermöglichen und den User ein größere Interaktivität und einen großen Bewegungsspielraum in einer Szene zu gewähren.

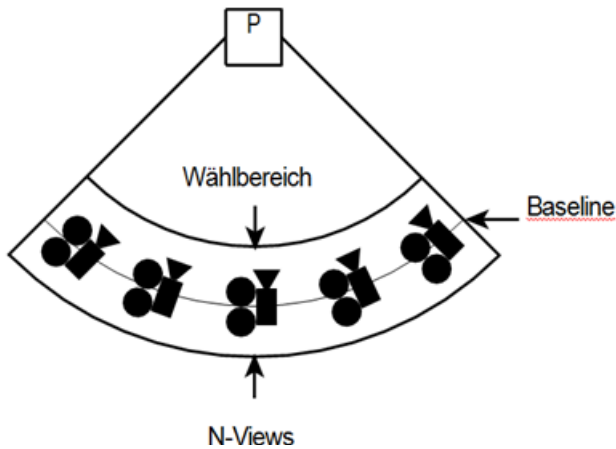


Abbildung 3.1.: Free Viewpoint Video

3.3.1.1. Model-based Free Viewpoint Video

Model-based Free Viewpoint Video ist ein Verfahren, welches im Idealfall die freie Navigation innerhalb der Bildszene erlaubt. Dies ist möglich, indem in Echtzeit aus den einzelnen Kamerabildern ein dreidimensionales Computermodell der Szene er-

²² Vgl. Ozaktas/Onural, Three-Dimensional Television, (2008), S. 320f

²³ Vgl. Eisert et al., Free View Point Video, (2006), S. 1.

stellt wird²⁴. Model-based Free Viewpoint Video benötigt zahlreiche komplexe Algorithmen zur Bilderkennung und ist rechenintensiv und gleichzeitig fehleranfällig. Die Qualität der Rekonstruktion kann aber durch Übertragung der Kameraspezifikationen und Optimierung der Randbedingung, wie Szenenausleuchtung verbessert werden²⁵. Trotzdem kann das Verfahren im besten Fall nur eine vereinfachte Näherung der realen Szene erstellen.

3.3.1.2. Depth-based Free Viewpoint Video

Depth-based Free Viewpoint Video wird auch als Depth Image Based Rendering (DIBR) bezeichnet und bezieht Depth-Maps bei der Berechnung von Zwischenbildern ein. Dieses Verfahren ist nicht geeignet die komplette Szene aus den einzelnen Kamerabildern zu rekonstruieren und so beliebige Kamerafahrten zu ermöglichen. Stattdessen ermöglicht dieses Verfahren die freie Navigation innerhalb eines Bereiches in der Nähe der Kameraachsen, der sogenannten Baseline²⁶. Dazu wird durch Depth Estimation von Kamerabildern eine Tiefenkarte für jedes Kamerabild erstellt. Aus diesen Tiefenkarten können dann beliebige virtuelle Views auf der Baseline zwischen zwei Kamerabildern interpoliert werden.

Die Generierung einer genauen Tiefenkarte für jedes Kamerabild ist bei diesem Verfahren daher äußerst wichtig. Um dies sicherzustellen werden jeweils für zwei benachbarte Kamerabilder zwei Disparitätskarten erstellt, die beide jeweils die Unterschiede zwischen beiden Bildern enthalten. Für mehrere Kameras entstehen so mehrere Disparitätskarten, die anschliessend gegeneinander verglichen- und zu einer einzigen Tiefenkarte zusammengefasst werden²⁷. Das exakte Vorgehen ist in dem Artikel von Eric Schumacher zum Thema Tiefenkarten in diesem Arbeitspapier zu finden.

3.3.1.3. Image-based Free Viewpoint Video

Image-based Free Viewpoint Video ist ein rein bildbasierendes System bei dem keine dreidimensionale Rekonstruktion der Szene erfolgt. Stattdessen werden aus den realen Kamerabildern durch digitale Bildverarbeitung Zwischenbilder erzeugt²⁸. Der Nachteil dieses Verfahren ist, dass einerseits hochauflösende Kamerabilder vorliegen müssen, andererseits muss auch eine große Zahl an Kameras vorhanden sein. Auch ist der Bereich in dem sich der Benutzer virtuell frei bewegen kann, ist stark eingeschränkt²⁹.

24 Vgl. Ozaktas/Onural, Three-Dimensional Television, (2008), S. 320f.

25 Eisert et al., Free View Point Video, (2006), S. 7f.

26 Eisert et al., Free View Point Video, (2006), S. 5f.

27 Eisert et al., Free View Point Video, (2006), S. 6f.

28 Vgl. Ozaktas/Onural, Three-Dimensional Television, (2008), S. 321f.

29 Vgl. Eisert et al., Free View Point Video, (2006), S. 6.

3.3.2. 3D-Video

Der zweite Bereich 3D-Video beschreibt das Anwendungsszenario des stereoskopischen Videos. Allerdings ist stereoskopisches Multi-view-Video nicht streng auf die effiziente Codierung und Decodierung zweier Blickwinkel beschränkt. Stattdessen können beliebige zusätzliche Blickwinkel übertragen oder durch Techniken wie Depth Image Based Rendering in Echtzeit erstellt werden. Dadurch können bei stereoskopischen Video die Bewegungsparallaxe und Okklusionseffekte realistisch dargestellt werden, wenn das Abbildungssystem z.B. ein autostereoskopisches Display ein Trackingfunktion besitzt. Mit solch einem System kann die relative Position des Zuschauers zum Abbildungssystem erfasst werden und parallel zur Bewegung des Zuschauers zwischen einzelnen Views umgeschaltet werden³⁰.

3.4. MVC-Architektur

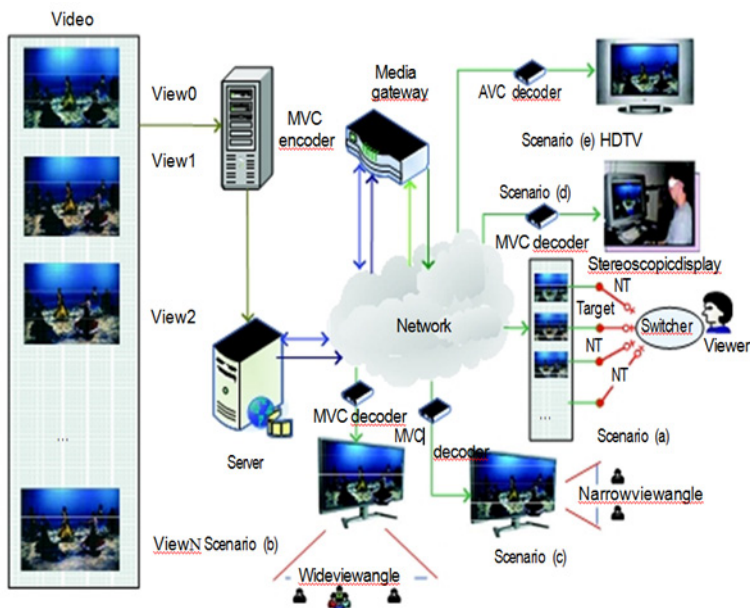


Abbildung 3.2.: MVC-Architektur

Ein MVC-System besteht aus drei Teilbereichen. In ersten Bereich befindet sich alle Multiview Videodateien bzw. Blickwinkel. Jeder Blickwinkel beinhaltet das aufge-

³⁰ Vgl. Ozaktas/Onural, Three-Dimensional Television, (2008), S. 318f.

nommene Videosignal einer Kamera. Alle Blickwinkel werden vom Multi-view Video Encoder zu einem einzelnen Videostrom (MVC-Bitstreamen) encodiert.

Im zweiten Teilbereich wird dieser Videostrom über ein Trägermedium und gegebenenfalls das Media Gateway an verschiedene Empfänger (Clients) weitergeleitet. Das Trägermedium muss nicht zwangsläufig das Internet sein. Auch Datenträger wie die Bluray-Disc sind geeignet, um Multi-view Video zu übertragen. Das Media Gateway wird auch als Mediaaware Network Element (Mane) bezeichnet. Es ist in der Lage eingehende Videosignale zu verarbeiten, benötigte Blickwinkel aus dem Videostrom zu extrahieren und nur die notwendigen Videoinformationen an bestimmte Empfänger zu verteilen. Für die Kommunikation zwischen Media Gateway und dem Empfänger ist ein rückkanalfähiges Medium notwendig.

Im dritten Teilbereich wird der Videostrom vom Multi-view Video Encoder für das Zielsystem des Empfängers decodiert und ausgegeben. Abhängig vom Ausgabesystem muss gegebenenfalls noch eine weitere Verarbeitung der Videodateien durchgeführt werden z.B. Erzeugung von Zwischenbildern bei Free Viewpoint Video³¹.

3.5. Anforderungen an MVC

3.5.1. Codiereffizienz

Da bei Multi-view Video zahlreiche Blickwinkel in einem Videostrom übertragen werden, steigt die Übertragungsmenge und Übertragungsdauer stark an. Daher muss eine effizientere Möglichkeit zur Steigerung der Codiereffizienz und Reduktion der übertragenden Datenmengen gefunden werden ohne dass sich die Komplexität der Codierung zu stark erhöht³².

Durch Erweiterung der Codierung auf Abhängigkeiten zwischen Blickwinkeln eines Zeitpunkts (die sogenannte Interview Prediktion) kann die Datenrate im Vergleich zur autonomen Parallelübertragung deutlich gesenkt werden³³.

3.5.2. Wahlfreier Zugriff

Multi-view Video benötigt vorallem im Rahmen von Free Viewpoint Video einen wahlfreien Zugriff auf Frames des selben Blickwinkels (Temporale Random Access) als auch Frames eines beliebigen Blickwinkels (View Random Access), ohne dass zu viele Frames decodiert werden müssen. Auch müssen mehrere Blickwinkel parallel verarbeitet werden können³⁴.

31 Vgl. Ying Chen et al., The Emerging MVC Standard, (2008), S. 01f.

32 Vgl. Ying Chen et al., The Emerging MVC Standard, (2008), S. 03.

33 Vgl. Ozaktas/Onural, Three-Dimensional Television, (2008), S. 326.

34 Vgl. Ozaktas/Onural, Three-Dimensional Television, (2008), S. 328.

3.5.3. Skalierbarkeit

Ein Videostrom wird als skalierbar bezeichnet, wenn man einen Informationsteil aus diesem extrahieren kann und beide Videoströme trotzdem valide und abspielbar bleiben. Die Erzeugung eines neuen Videostroms mit geringerer Framerate wird als temporal Scalability bezeichnet. Die Erzeugung eines neuen Videoteilstroms mit geringerer Auflösung gegenüber dem Originalvideostrom nennt man partial Scalability (Bildgröße). Multi-view Video benötigt eine zusätzlich Art von Scalability in Bezug zur Extraktion von Blickwinkeln aus einem Videostrom. Diese sogenannte View Scalability, sagt aus, dass aus einem Videostrom Blickwinkel entfernt werden können und der Videostrom trotzdem valide und abspielbar bleibt³⁵. Besonders für Free Viewpoint Video und der Verteilung von MVC-Videoströmen durch Mane ist dies von besonderer Wichtigkeit.

3.5.4. Abwärtskompatibilität

Multi-view Video sollte zu bestehenden H264/AVC-Decodern kompatibel sein, so dass mindestens ein Videostrom decodiert werden kann. Dies wird durch die Einbettung eines sogenannten Baseviews erreicht, welcher in H264/AVC codiert ist und so keine Abhängigkeiten zu anderen Blickwinkeln besitzt³⁶.

3.5.5. Übertragung von Metadaten

In die Erstellung von virtuellen Zwischenbildern können die Kameraparameter einbezogen werden, um die virtuelle Rekonstruktion der Szene zu verbessern. Daher sollte jeder View die Kameraparameter als Metadaten beinhalten und übertragen³⁷.

35 Ozaktas/Onural, Three-Dimensional Television, (2008), S. 328.

36 Ozaktas/Onural, Three-Dimensional Television, (2008), S. 328.

37 Vgl. Ozaktas/Onural, Three-Dimensional Television, (2008), S.328

4. Stereoskopische Videocodierung mit H264/MVC

4.1. Allgemein

H264/MVC ist eine Erweiterung des H264/AVC Videostandards und baut daher auf dessen grundlegenden Codiertechniken auf und erweitert diese in einigen Bereichen, wie z.B. Aufbau des Videostroms, um eine effizientere Codierung von Multi-view Video zu ermöglichen.

Ein H264/MVC Encoder unterscheidet sich von einem H264/AVC Encoder zunächst in der Größe des Decoded Picture Buffer (DPB), da dieser für Multi-view Video erhöht werden musste. Aufgrund des höheren DPB müssen für H264/MVC neue Profile definiert werden mit angepasster Größe des DPB³⁸.

4.2. Videocodierung mit H264/AVC

Videocodierung mit H264/AVC erfolgt in vier Teilschritten. Zuerst wird der aktuelle Videoframe (F) in mehrere Slices unterteilt, die wiederum Makroblöcke enthalten³⁹. In jedem Makroblock sind die Luminanz und Chrominanzwerte gespeichert. Aufgrund des Chromasubsamplings liegen die Luminanzwerte nur mit halbiertem Auflösung vor. Die Größe der Makroblöcke ist auf 16x16 Pixel festgelegt. Makroblöcke können wiederum in verschiedene Makroblockpartitionen unterteilt werden z.B. 8 x 8 Pixel Blöcke. Diese Makroblockpartitionen können selbst wieder in Sub-Makroblockpartitionen von beispielsweise 4 x 4 Pixel unterteilt werden. Diese so entstehende Unterteilung wird als Tree Structured Motion Compensation bezeichnet⁴⁰.

Neben der Makroblockgröße unterscheidet man auch den Typ eines Makroblocks und damit die Art der Slices. I-Makroblöcke werden von vorherigen Makroblöcken des selben Slices codiert und dienen anderen Makroblöcke als Referenz. P-Makroblöcke referenzieren auf vorhergegangene I-Makroblöcke von vorherigen Frames und ermöglichen dadurch eine Steigerung der Kompressionsrate. B-Makroblöcke referenzieren sowohl auf vorhergegangene- als auch auf nachfolgende I-Makroblöcken dieser Frames und ermöglichen so eine weitere Steigerung der Kompression⁴¹.

Analog zu den Makroblöcken werden auch die Slices nach I, P und B-Slice unterschieden. I-Slices enthalten nur I-Makroblöcke, während P-Slices sowohl I-Makroblöcke als auch P-Makroblöcke enthalten können. B-Slices können I-Makroblöcke und B-Makroblöcke enthalten⁴². Zur effizienteren Verwaltung der Referenzbilder bei

38 Vgl. Ozaktas/Onural, Three-Dimensional Television, (2008), S. 323f.

39 Vgl. Hottong, Digitale AV-Technik, (2009), S. 146.

40 Vgl. Richardson, H.264 and MPEG-4 Video Compression, (2008), S. 160.

41 Vgl. Richardson, H.264 and MPEG-4 Video Compression, (2008), S. 164f.

42 Vgl. Richardson, H.264 and MPEG-4 Video Compression, (2008), S. 159f

Verwendung von B-Slices werden diese in einen Index eingetragen.

Bei der Codierung eines Frames (F) wird für jeden Makroblock eine Voraussage getroffen. Diese kann eine intra-basierte Voraussage sein, die auf der Codierung von vorangegangener Makroblöcken des gleichen Slice beruht. Alternativ kann es auch eine intrabasierte Voraussage sein, die auf der Referenzierung und Codierung von Makroblöcken aus vorausgegangenen- und nachfolgenden Frames beruht. In beiden Fällen wird die Differenz zwischen dem aktuellen Makroblock und der Voraussage gebildet und die Abweichungen zwischen beiden Makroblöcken gespeichert⁴³.

Im nächsten Schritt werden diese Abweichungen quantisiert und die Koeffizienten in einem Zickzack-Verfahren angeordnet⁴⁴. Dieses Ausleseverfahren ist für die anschließende Entropiecodierung besser geeignet, da so ähnliche Koeffizienten nacheinander stehen und diese somit einfacher zusammenfassbar sind⁴⁵. Bei H264/AVC kommen mehrere Verfahren zur Entropiecodierung zum Einsatz. Neben Variable-Length Codes (VLC) auch Context-Adaptive Arithmetic Coding (CABAC), Context-Adaptive Variable Coding (CAVLC) und Exp-Golomb⁴⁶.

Im letzten Schritt werden die entropiecodierten Koeffizienten zu einem Videostrom zusammengefasst und durch den Network Abstraction Layer (NAL) übertragen⁴⁷.

4.2.1. Aufbau eines H264/AVC-Videostroms

Zur einfacheren Übertragung H264/AVC-codierter Videodaten über packetbasierte Übertragungsmedien wie beispielsweise das Internet, werden Videodaten nach H264/AVC standardmäßig in Network Abstraction Layer (NAL) und Video Coding Layer (VCL) unterteilt. Ersterer beinhalten Informationen die für den Transport des Videostrom wichtig sind, während im Video Coding Layer die eigentlichen encodierten Daten gespeichert werden. Alle Videodaten werden zur Übertragung in NAL-Units gekapselt. Diese bestehen aus einem 1 byte großem NAL-Header und dem Raw Byte Sequence Payload (RBSP) beliebiger Größe. NAL-Units, die Slices im RBSP beinhalten, werden VCL NAL-Unit genannt⁴⁸.

Im NAL-Header einer NAL-Unit wird die Art des RBSP definiert. Bei einer nicht VCL NAL-Unit kann dies ein ein Sequence Parameter Set sein, welches globale Informationen über eine Videosequenz enthält, wie beispielsweise Höhe und Breite des Videobildes.

43 Vgl. Richardson, H.264 and MPEG-4 Video Compression, (2008), S. 161f

44 Vgl. Richardson, H.264 and MPEG-4 Video Compression, (2008), S. 198f

45 Vgl. Hottong, Digitale AV-Technik, (2009), S. 130.

46 Vgl. Richardson, H.264 and MPEG-4 Video Compression, (2008), S. 198f.

47 Vgl. Richardson, H.264 and MPEG-4 Video Compression, (2008), S. 161f

48 Vgl. Richardson, H.264 and MPEG-4 Video Compression, (2008), S. 163f.

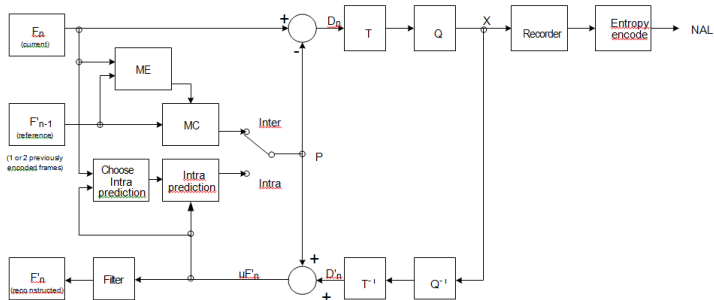


Abbildung 4.1.: H264/AVC-Encoder

Weiterhin kann es ein Picture Parameter Set sein, welches Informationen für einzelne Frames enthält, z.B. Anzahl an referenzierten Frames etc. Diese Informationen werden nur jeweils einmalig übertragen und bleiben für die Videosequenz gültig, bis eine NAL-Unit mit neuen Parametern übermittelt wird. So kann eine redundante Übertragung dieser Parameter vermieden werden und dadurch die Kompression erhöht werden⁴⁹.

Im Falle von VCL NAL-Units sind es die eigentlichen encodierten Videodaten⁵⁰.

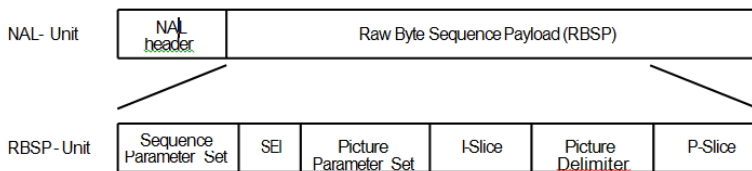


Abbildung 4.2.: Aufbau einer H264/AVC NAL-Unit

4.3. MVC-Erweiterungen gegenüber H264/AVC

4.3.1. Interview Prediktion

Zur Steigerung der Effizienz der Codierung von Multi-view Video wird neben der temporalen Interframecodierung, welche die Redundanz zwischen zeitlich aufeinanderfolgenden Frames zur Datenreduktion nutzt, zusätzlich Interviewcodierung angewendet, welche die Redundanz zwischen benachbarten Blickwinkeln bei der Codierung einbezieht. Dies wird als Interview Prediktion bezeichnet. Die ausschließliche Codierung jedes Blickwinkels mit temporaler Interframecodierung wird als Simulcast bezeichnet⁵¹.

49 Vgl. Ying Chen et al., *The Emerging MVC Standard*, (2008), S. 04f.

50 Vgl. Richardson, H.264 and MPEG-4 Video Compression, (2008), S. 220f

51 Vgl. Merkle et al., Efficient Prediction Structures for Multi-view Video Coding, (2006), S. 03.

Die Codierung mit Interview Prediktion erfolgt anhand einer flexiblen Prediktionsstruktur, die auf hierarchischen B-Frames basiert. Die Interview Prediktion kann sowohl nur auf intra-codierte Keyframes als auch auf alle Arten von Frames innerhalb einer GOP (Group of Picture) angewendet werden. Bei erster Variante werden auf dem ausschließlich interframe-codierten Baseview aufbauend, dessen I-Frames als Referenz für die intercodierten P-Frames in jedem Blickwinkel benutzt. Die restlichen Frames der GOP jedes Views werden weiterhin mit Interframecodierung encodiert.

Zur Steigerung der Codiereffizienz kann die Interview Prediktion auch auf alle Arten von Frames eines Views ausgeweitet werden. Die Interview Prediktion zwischen den Keyframes mehrerer Views und bleibt dabei unverändert. Zusätzlich werden aber die intracodierten Frames innerhalb eines Views von anderen intra-codierten Frames eines anderen Views referenziert. Dadurch steigt neben der Codiereffizienz allerdings auch die Komplexität der Codierung stark an⁵².

Statistisch gesehen ist Interview Prediktion am effektivsten bei Videoströmen mit geringen Frameraten und Kameras, die sehr nahe beieinander positioniert sind, also z.B. auch Stereomaterial mit moderaten Stereobasen (Interocular distance). Besonders bei Szenen mit viel Bewegungen im Videobild erzielen bei der Codierung mit Interview Prediktion eine effektive Kompressionsrate, während statische Bildinhalte effektiver durch temporalen Interframecodierung codiert werden kann⁵³.

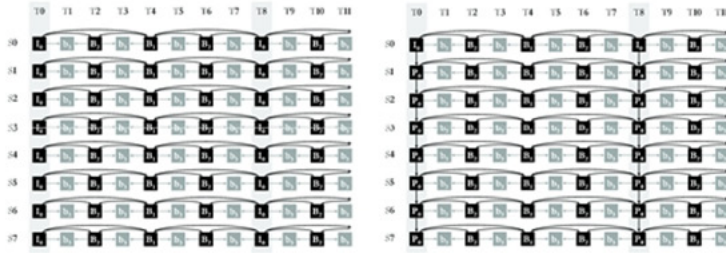
4.3.2. Aufbau eines MVC-Videostroms

Um die Abwärtskompatibilität mit regulären H264/AVC-Decodern zu gewährleisten, wird in MVC-Videoströmen ein Baseview eingebettet. Dieser besteht bis auf Prefix-NAL-Units ausschliesslich aus regulären H264/AVC NAL-Units. Prefix-NAL-Units sind jeder VCL NAL-Unit des Baseviews vorangestellt und enthalten zusätzliche Codierungsinformationen für weitere Blickwinkel. Da Prefix-NAL-Units von H264/AVC-basierenden Decoder nicht decodiert werden, bleibt die Abwärtskompatibilität mit regulären H264/AVC-Decodern gegeben⁵⁴.

52 Vgl. Merkle et al., Efficient Prediction Structures for Multi-view Video Coding, (2006), S. 03.

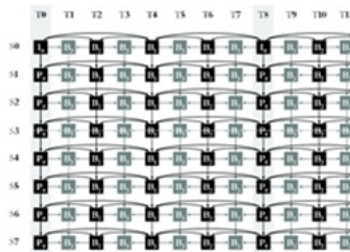
53 Vgl. Ozaktas/Onural, Three-Dimensional Television, (2008), S. 323f.

54 Vgl. Ying Chen et al., The Emerging MVC Standard, (2008), S. 04f.



Temporalen Interframecodierung

Interview Prediction (Keyframes)



Interview Prediction

Abbildung 4.3.: Interview Prediction

Zur Codierung und Übertragung von Multi-view Video wird die zusätzliche VCL NALUnit Coded Slice of MVC eingeführt. In dieser NAL-Unit werden nur encodierte Videodaten von anderen Blickwinkeln gespeichert. Sie bestehen aus einem 4 bit großem Header und der beliebig großen Raw Byte Sequence Payload. Im gegenüber regulären NAL-Units vergrößertem Header werden zusätzliche Parameter wie die Nummer der Blickwinkels. Dies erleichtert den Zugriff auf einzelne Views bei der Decodierung. MVC-codierte NAL-Units werden von H264/AVC-Decodern nicht decodiert⁵⁵.

Zusätzlich werden die Abhängigkeiten zwischen einzelnen Views in den Sequence Parameter Set von regulären NAL-Units gespeichert. Dadurch kann auf diese, für Interviewprediction notwendige Informationen, schneller zugegriffen werden⁵⁶.

⁵⁵ Vgl. Ying Chen et al., The Emerging MVC Standard, (2008), S. 04f.

⁵⁶ Vgl. Ying Chen et al., The Emerging MVC Standard, (2008), S. 04f.

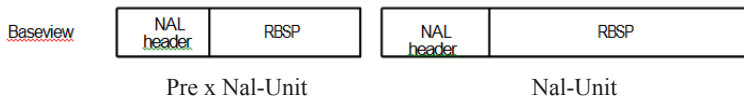


Abbildung 4.4.: Aufbau NAL-Unit eines Baseview

4.3.3. Zugriff und Anpassung eines MVC-Videostroms

MVC benutzt zeitliche und räumlich flexible Skalierbarkeit. Das bedeutet, dass sich Teile eines MVC-Videostroms auf eine bestimmte Framerate und ein bestimmten View beziehen.

Es werden gleichzeitig mehrere Videostrome mit unterschiedlichen Views und Frameraten im Videovertersystem Mane vorgehalten und nur benötigte Views und Frameraten an den Client übertragen. Nicht benötigte Informationen werden nicht an den Client übertragen.

Eine NAL-Unit eines Videostroms zu einem festen Zeitpunkt t steht für einen bestimmte Zeitpunkt und View. Dies wird als Operation Point (op) bezeichnet. Beide Parameter werden in einer Baumstruktur angeordnet. Im erweiterten Header jeder MVC-codierten NAL-Unit werden die zwei Werte View id und Temporal id vorhanden. View id speichert die Nummer des aktuellen Views, während Temporal id, die Framerate speichert. Operation Points mit einer hohen Temporal id beziehen sich auf NAL-Units mit einer hohen Framerate während eine kleine Temporal id auf Nal-Units einer geringe Framerate verweist. Bei der Differenzbildcodierung referenzieren NAL-Units mit hoher Temporal ID nur auf NAL-Units mit geringer Temporal ID.

Beide Parameter ermöglichen die Extraktion von Teilen eines Videostroms ohne das komplette Parsen des gesamten Videostroms. Dies ist besonders bei Free Viewpoint Video wichtig. Die Verknüpfung eines Operationpoint mit den benötigten NAL-Units ist in der VSSEI (View Scalability Information SEI Message) gespeichert. Wenn ein bestimmter Operationpoint und damit auch ein bestimmter View und Framerate angesprochen werden soll, werden die damit verbundenen NAL-Units vom Videovertersystem extrahiert und die anderen nicht benötigten NAL-Units verworfen. Anschließend werden diese als skalierbarer Videostrom an den Client gesendet⁵⁷.

4.3.4. Speicherung von MVC-Videostromen

Die Reihenfolge in der NAL-Units innerhalb des Videostroms übertragen und gespeichert werden, bezeichnet man als Decoding Order. Da MVC-Videostrome mehrere Views in unterschiedliche Zeitenabfolgen besitzen, gibt es mehrere Möglichkeiten zur

⁵⁷ Vgl. Ying Chen et al., The Emerging MVC Standard, (2008), S. 05f.

Speicherung der NAL-Units. Standardmäßig sind in MVC zwei Varianten definiert⁵⁸.

4.3.4.1. View First Coding

View First Coding bedeutet, dass jeder Frame eines Views komplett nacheinander gespeichert wird. Wenn eine GOP bzw. alle Frames eines Views gespeichert sind, wird die GOP des nächsten Views gespeichert. Das bedeutet, dass der Decoder die Decodierung eines neuen Views erst beginnen kann, wenn alle Bilder des vorherigen Views decodiert oder durchlaufen worden sind. Weiterhin birgt View First Coding den Nachteil, dass NAL-Units verschiedener Views des gleichen Zeitpunkt nicht zusammen gespeichert werden. Stattdessen werden NAL-Units verschiedener Zeitpunkte nacheinander gespeichert, was bei einem wahlfreien Zugriff und Extraktion der NAL-Units zu einem höheren Delay führt⁵⁹.

4.3.4.2. Time First Coding

Time First Coding bedeutet, dass die Frames verschiedener Views eines Zeitpunkts nacheinander gespeichert werden. Alle Views eines Zeitpunkts bilden eine Einheit und der Decoder beginnt die Decodierung einer neuen Einheit, wenn alle Frames der aktuellen Einheit decodiert oder durchlaufen worden sind. Dadurch sind alle NAL-Units in korrekter fortlaufender Reihenfolge⁶⁰.

58 Vgl. Ying Chen et al., The Emerging MVC Standard, (2008), S. 07f

59 Vgl. Ying Chen et al., The Emerging MVC Standard, (2008), S. 07f.

60 Vgl. Ying Chen et al., The Emerging MVC Standard, (2008), S. 08f.

5. Der JMVC-Codec in der Praxis

5.1. Überblick

Zu einer ersten Evaluation der MVC-Erweiterung für den Einsatz in der Distribution von S3D-Filmen wird im Folgenden der JMVC-Codec (Joint Multiview Video Coding) verwendet. Dieser ist der Referenzcodec des Multiview Video Coding (MVC) Projekts des Joint Video Teams (JVT) und der ISO/IEC Moving Pictures Experts Group (MPEG) und der ITU-T Video Coding Experts Group (VCEG). Neben dem Codec sind im JMVC-Softwarepakets auch Decoder, Multiplexer und Testprogramme enthalten.

Das JMVC-Softwarepaket ist im Quelltext frei verfügbar. Der Quelltext für das Softwarepaket ist in C++ geschrieben und z.B. auf dem Webspace der Rheinisch-Westfälische Technische Hochschule Aachen (RWTH-Aachen) sowohl in einem CVS-Repository⁶¹ als auch als ZIP-Datei⁶² verfügbar. Allerdings werden aktuelle Versionen des Codecs nur noch durch das CVS verbreitet. Um diese Abzurufen ist eine geeignete CVS-Software wie Tortoise⁶³ notwendig⁶⁴.

Unsere Untersuchungen wurden mit der Version JMVC 7 0 vom 14. Dezember 2009 durchgeführt⁶⁵.

5.2. Installation und Kompilierung

5.2.1. CVS-Zugangsdaten

Authentifizierungstyp	pserver
Hostadresse	garcon.ient.rwth-aachen.de /cvs/jvt
CVS-Pfad	jvtuser
Benutzername	jvt.Amd.2
Passwort	jmvc
Modulname	

Tabelle 5.1.: Zugangsdaten für das JMVC CVS-Repository der RWTH Aachen

61 CVS (Concurrent Versions System) ist ein weitverbreitetes System zur Versionsverwaltung von Softwarequelltexten

62 http://wftp3.itu.int/av-arch/jvt-site/2009_01_Geneva/JVT-AD207.zip

63 <http://www.tortoisecvs.org>

64 Vgl. Joint Video Team, MVC Software Manual, (2009), S. 03f.

65 Stand 17.02.2010

5.3. Bestandteile des JMVC-Softwarepaket

5.3.1. H264AVCEncoderLibTestStatic

Dieses Programm ist der eigentliche MVC-Encoder mit welchem unkomprimierte YUV-Videodateien in MVC-Videoströme encodiert werden. Der Encoder wird mit folgendem Aufruf gestartet:

H264AVCEncoderLibTestStatic -vf [Konfigurationsdatei.cfg] [View]

Der Parameter -vf verweist auf die eine Textdatei, in welcher alle Konfigurationseinträge zentral gesammelt werden. Der zweite Parameter View gibt die Nummer des aktuellen Views an. Die Zähler beginnt bei 0, daher ist View 0 der erste View. Der Encoder wird für jeden View nacheinander aufgerufen. Weiterhin müssen bei eingeschalteter Interviewprediction alle vorhergehenden Views im Verzeichnis bereits encodiert vorliegen⁶⁶.

In der derzeitigen Version des MVC-Encoders gibt es keine Möglichkeit die Bitrate für den Encodierdurchgang direkt festzulegen. Stattdessen wird die Codierung über den Quantisierungsparameter bestimmt. Der Encoder variiert den Quantisierungsparameter je nach Art des zu codierenden Frames. I-Frames haben z.B. einen niedrigeren Quantisierungsparameter als B-Frames.

The image contains two screenshots of a Windows command prompt window running the H264AVCEncoderLibTestStatic program. The top screenshot shows the initialization for View 0 (config.cfg 0) and the encoding of the first frame (I-frame) with a quantization parameter of 5. The bottom screenshot shows the initialization for View 1 (config.cfg 1) and the encoding of the first frame (I-frame) with a quantization parameter of 5. Both screenshots show the program's output, including the initialization parameters for NAL units, the encoding of the first frame, and the preparation to remove multiview references.

```

C:\Windows\system32\cmd.exe - H264AVCEncoderLibTestStatic -vf config.cfg 0
SPS: num_non_anchor_refs: 1(1): 0
Initializing parameters for NAL_UNIT_SPS...
h264SPS=1 NumRef=1 Max_NumRefFrames=5 encDPBsize=0 decDPBsize=5 LevelIdc=31
ViewCodingOrder: 0 1
Initializing parameters for NAL_UNIT_SURSEQI_SPS...
h264SPS=0 NumRef=2 Max_NumRefFrames=5 encDPBsize=12 decDPBsize=11 LevelIdc=31
ViewCodingOrder: 0 1
0 I IDR REF_0IEM 0 QP 5 Y 50.5093 dB U 56.5352 dB U 56.5739 dB bits
3876360
Preparing to remove multiview refs
1 P REF 4 QP 5 Y 50.3695 dB U 56.5480 dB U 56.5654 dB bits
3828832
Preparing to remove multiview refs
2 B REF 2 QP 0 Y 51.0725 dB U 51.7422 dB U 51.7650 dB bits
1924800
Preparing to remove multiview refs
3 B 1 QP 9 Y 51.2082 dB U 51.2700 dB U 51.2820 dB bits
1731600
Preparing to remove multiview refs

C:\Windows\system32\cmd.exe - H264AVCEncoderLibTestStatic -vf config.cfg 1
Initializing parameters for NAL_UNIT_SURSEQI_SPS...
h264SPS=0 NumRef=2 Max_NumRefFrames=5 encDPBsize=12 decDPBsize=11 LevelIdc=31
ViewCodingOrder: 0 1
Preparing to add multiview refs
going to read frame 0 for multiview ref
hidden frame with frameNumber and poc (0, 0)
0 P IDR 0 QP 5 Y 50.4526 dB U 56.5268 dB U 56.5540 dB bits
3881856
Preparing to remove multiview refs
going to remove frame with frameNumber and poc (0, 0)
Preparing to add multiview refs
1 P REF 4 QP 5 Y 50.3812 dB U 56.5571 dB U 56.5437 dB bits
3854704
Preparing to remove multiview refs
Preparing to add multiview refs
2 B REF 2 QP 0 Y 51.0479 dB U 51.7160 dB U 51.7210 dB bits
1907296
Preparing to remove multiview refs

```

Abbildung 5.1.: Aufruf des MVC-Encoders für View 0 (oben) und View 1 (unten)

⁶⁶ Vgl. Joint Video Team, MVC Software Manual, (2009), S. 07f.

5.3.1.1. Wichtige Parameter der Konfigurationsdatei

InputFile	Dateinamen der zu komprimierenden Dateien ohne Dateiendung. Die Dateien sollten nach folgendem Muster aufgebaut sein: „InputFile_view“, z.B. „sourcefile_0, sourcefile_1 usw.“
OutputFile	Dateiname der zugenerierenden Bitsreamdateien. Während des Comprimierens werden die Dateien automatisch nach folgendem Muster angelegt: „Outputfile_0, Outputfile_1 usw.“
ReconFile	Dateiname der automatisch generierten Ausgangsdateien. Sie dienen für Debugg-Prozesse Während des Comprimierens werden die Dateien automatisch nach folgendem Muster angelegt: „Reconfile_0, reconfile_1 usw.“
SourceWidth	Definiert die Breite der Quellevideodatei in Pixel. Die Breite muss ein Vielfaches der Zahl 16 sein.
SourceHeight	Definiert die Höhe der Quellevideodatei in Pixel. Die Höhe muss ein Vielfaches der Zahl 16 sein.
FrameRate	Definiert die Framerate der Quellevideodatei
FramesToBeEncoded	Definiert die Anzahl der Frames die pro Enkodierdurchgang enkodiert werden sollen.
BasisQP	Definiert den Quantisierungsparameter für den Enkodiervorgang.

Tabelle 5.2.: Übersicht über wichtigsten Parameter der Konfigurationsdatei

5.3.1.2. Interview Prediction aktivieren und deaktivieren

Die Interviewprediction zwischen einzelnen Views wird ebenfalls in der globalen Konfigurationsdatei konfiguriert. Zuerst wird NumViewsMinusOne auf die Anzahl der Views minus 1 gesetzt. Bei zwei Views muss dieser Wert auf 1 gesetzt werden. Danach wird die Reihenfolge der Views mit ViewOrder festgelegt.

NumViewsMinusOne 1

ViewOrder 0-1

Anschliessend wird für jeden View seperat definiert, auf welche vorherigen und nachfolgenden Views diese referenzieren sollen. View 0 ist der Baseview und dieser referenziert auf keinen anderen View, daher ist der Parameter Fwd_NumAnchorRefs mit 0 belegt und Fwd_AnchorRefs nicht vorhanden.

View 1 dagegen referenziert bei eingeschalteter Interviewprediction auf View 0, daher muss für View 1 Fwd_NumAnchorRefs mit 1 gesetzt sein und Fwd_AnchorRefs mit 0 0 belegt

sein. Der Parameter definiert, dass View 1 auf View 0 referenziert. Bei deaktivierter Interviewprediction sind analog zu View 0 beide Parameter nicht belegt bzw. vorhanden⁶⁷.

Konfiguration von View 0 und View 1 bei eingeschalteter Interview Prediktion

```
View_ID 0
Fwd_NumAnchorRefs      0
Bwd_NumAnchorRefs      0
Fwd_NumNonAnchorRefs 0
Bwd_NumNonAnchorRefs  0
```

```
View_ID 1
Fwd_NumAnchorRefs      1
Bwd_NumAnchorRefs      0
Fwd_NumNonAnchorRef    0
Bwd_NumNonAnchorRefs   0
Fwd_AnchorRefs  0 0
```

Konfiguration von View 0 und View 1 bei ausgeschalteter Interview Prediktion

```
View_ID 0
Fwd_NumAnchorRefs      0
Bwd_NumAnchorRefs      0
Fwd_NumNonAnchorRefs   0
Bwd_NumNonAnchorRefs   0
```

```
View_ID 1
Fwd_NumAnchorRefs      0
Bwd_NumAnchorRefs      0
Fwd_NumNonAnchorRef    0
Bwd_NumNonAnchorRefs   0
```

5.3.2. MVCBitStreamAssembler

Dieses Programm ist der Multiplexer⁶⁸ für MVC-Videoströme. Er fügt die separat vorliegenden MVC-Videoströme jedes Views zu einem einzelnen MVC-Videostroms zusammen und entkapselt die Videodaten in NAL-Units⁶⁹. Der Multiplexer wird mit folgendem Aufruf gestartet:

⁶⁷ Vgl. Bal, Three-Dimensional Video Coding on Mobile, (2009), S. 55f.

⁶⁸ Als Multiplexer bezeichnet man in der Videotechnik ein Programm mit dem mehrere Videoströme in einer einzelnen Videostrom zusammengefügt werden.

⁶⁹ Vgl. Joint Video Team, MVC Software Manual, (2009), S. 17

MVCBitStreamAssemblerStatic -vf [Konfigurationsdatei.cfg]

Der Parameter -vf verweist auf die eine Textdatei, in welcher die Dateinamen der einzelnen MVC-codierte Views aufgelistet sind und deren Reihenfolge. Die Konfigurationsdatei besitzt folgenden Aufbau:

```

OutputFile      ausgabedatei.264
NumberOfViews   2
InputFile0      view0.264
InputFile1      view1.264
    
```

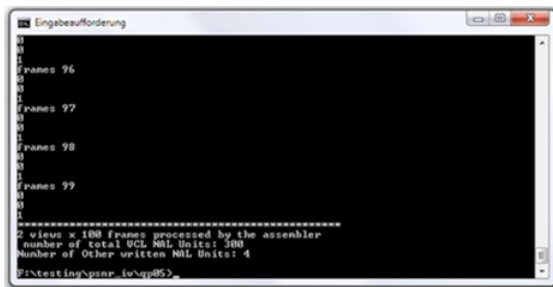


Abbildung 5.2.: Aufruf des MVCBitStreamAssemblers

5.3.3. H264AVCDecoderLibTestStatic

Dieses Programm ist der Decoder für MVC-Videoströme. Dieser decodiert eine zuvor durch das Programm MVCBitStreamAssembler erstellte MVC-codierte Videodatei, in welcher mehrere Views gespeichert sind. Anschließend extrahiert das Programm alle Views und wandelt diese jeweils in eine unkomprimierten YUV-Videodatei um⁷⁰. Der Decoder wird mit folgendem Aufruf gestartet:

H264AVCDecoderLibTestStatic [Bitstreamdatei] [Ausgabedatei] [Viewanzahl]

Der erste übergebene Parameter verweist auf die MVC-codierte Videodatei, die mehrere Views enthält. Der zweite Parameter bestimmt den Dateinamen unter welchem die einzelnen Views als YUV-Videodatei ausgegeben werden. Der Decoder fügt an den Dateinamen automatisch die Nummer des jeweiligen Views an z.B. yuvout 0.yuv. Der dritte Parameter gibt die Anzahl der Views an, die in der zu extrahierenden MVC-codierte Videodatei enthalten sind.

⁷⁰ Vgl. Joint Video Team, MVC Software Manual, (2009), S. 16.

```

C:\Windows\system32\cmd.exe
List 1: 1/74 1/76
C 0: Slice POC 95 < LId 0, TL X, QL 0, AUC-B, BId-1, AP 0, QP 10, FRAME >
List 0: 0/74 0/72
List 1: 0/74
C 1: Slice POC 95 < LId 0, TL X, QL 0, MUC-B, BId-1, AP 0, QP 10, FRAME >
List 0: 1/74 1/72
List 1: 1/74
C 0: Slice POC 98 < LId 0, TL X, QL 0, AUC-P, BId-1, AP 0, QP 9, FRAME >
List 0: 0/76
C 1: Slice POC 98 < LId 0, TL X, QL 0, MUC-P, BId-1, AP 0, QP 9, FRAME >
List 0: 1/76
C 0: Slice POC 97 < LId 0, TL X, QL 0, AUC-B, BId-1, AP 0, QP 10, FRAME >
List 0: 0/76
List 1: 0/78
C 1: Slice POC 97 < LId 0, TL X, QL 0, MUC-B, BId-1, AP 0, QP 10, FRAME >
List 0: 1/76
List 1: 1/78
C 0: Slice POC 99 < LId 0, TL X, QL 0, AUC-P, BId-1, AP 0, QP 10, FRAME >
List 0: 0/78 0/76
C 1: Slice POC 99 < LId 0, TL X, QL 0, MUC-P, BId-1, AP 0, QP 10, FRAME >
List 0: 1/78 1/76
200 frames decoded
F:\testing\psnr_io>H264AVCDecoderLibTestStatic assembled.264 source.yuv 2

```

Abbildung 5.3.: Aufruf des H264AVCDecoderLibTestStatic

5.3.4. PSNRStatic

Dieses Programm ist ein Testprogramm zur Bestimmung des Peak-Signal-To-Noise-Ratio (PSNR) zweier Videodateien. Weiterhin wird die Bitrate eines MVC-codierten Videostroms ausgegeben⁷¹. Das Programm wird mit folgendem Aufruf gestartet:

PSNRStatic [Breite] [Höhe] [Quelldatei] [Ausgabedatei]

Der erste übergebene Parameter bestimmt die Höhe beider Videodateien in Pixeln. Der zweite Parameter gibt die Breite in Pixeln an. Der dritte Parameter verweist auf die unkomprimierte Quelldatei. Der vierte Parameter verweist auf die zuvor encodierte Datei. Sowohl Quelldatei als auch die Ausgabedatei müssen im YUV-Format vorliegen. Daher muss die komprimierte Ausgabedatei vor der Durchführung der PSNR-Bewertung in eine YUV-Datei umgewandelt werden. Es ist zu empfehlen, dazu die beiden Programme MV-CBitStreamAssembler und H264AVCDecoderLibTestStatic zu benutzen, da andere Programme in der Regel MVC-codierte Videodateien nicht korrekt verarbeiten können. Die PSNR-Werte werden getrennt nach Y-PSNR (Luminanz), U-PSNR (Chrominanz U) und V (Chrominanz U) ausgegeben⁷².

```

Eingabeaufforderung
41.5352 45.4873 45.7431
41.6165 45.3859 45.6783
41.5450 45.4951 45.5471
41.6949 45.6305 45.7127
41.6289 45.6488 45.6839
41.7322 45.8036 45.8379
41.6323 45.6783 45.6246
42.3612 46.0770 46.1451
41.7438 45.5622 45.6477
41.8119 45.4448 45.7056
41.7477 45.5556 45.7118
41.5249 45.8041 45.4579
41.6672 45.6248 45.6841
41.7896 45.6868 45.4521
41.6883 45.8714 45.3674
42.3582 46.0748 46.1097
41.5556 45.8019 45.5715
41.5473 45.5244 45.4981
41.7436 45.3947 45.3778
total 41.8268 45.6318 45.8249
total 41.8268 45.6318 45.8249
F:\testing\psnr_io>PSNRStatic 1024 768 hallet_0.yuv compressed_0.yuv

```

Abbildung 5.4.: Aufruf des PSNRStatic Tools

⁷¹ Vgl. Joint Video Team, MVC Software Manual, (2009), S. 16f.

⁷² Vgl. Joint Video Team, MVC Software Manual, (2009), S. 17.

5.4. Testdateien

5.4.1. Stereoskopische Videodateien - Microsoft Research 3D Clips

Als erste stereoskope Testsequenzen diente der Testclip Ballet der Interactive Visual Media Group der Abteilung Microsoft Research 3D. Diese Testsequenz ist zur nicht-kommerziellen Nutzung freigeben und kann von der Website von microsoft heruntergeladen werden ⁷³.

Jede Testsequenz besteht aus einer kurzen Filmsequenz, die mit acht synchronisierten Kameras aufgenommen wurde. Zusätzlich ist für jeden View die Depthmap enthalten. Die Aufnahme erfolgte in 25 Hz mit einer Auflösung von 1024x768 Pixeln (SVGA). Die Testsequenzen liegen als MSI-Dateien vor und müssen nach dem Download installiert werden.

Nach erfolgreicher Installation sind die Testsequenzen im Ordner MSR3DVideo-Ballet im Programmverzeichnis von Windows zu finden. Jeder Frame liegt einzeln als Bitmapdatei vor.

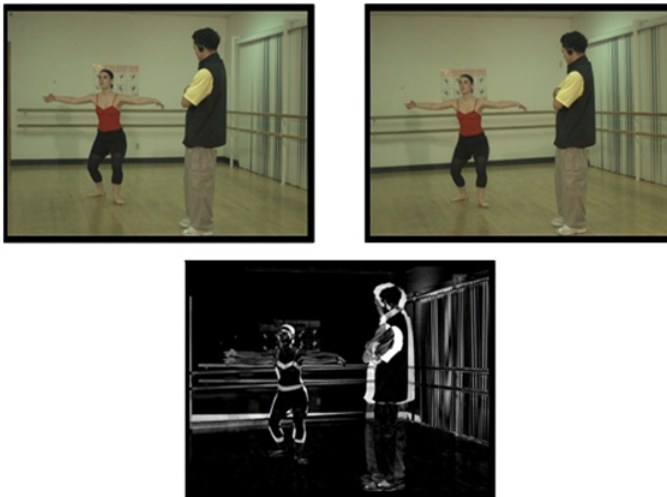


Abbildung 5.5.: Samples Microsoft Research 3D

5.4.2. MERL-Testsequenzen

Als weitere stereoskopische Testsequenz wird ein Testclip des Mitsubishi Electric Research Laboratories (MERL) verwendet. Dieser befinden sich öffentlich zugäng-

⁷³ <http://research.microsoft.com/en-us/um/people/sbkang/3dvideodownload/>

lich auf dem FTP Server des MERL⁷⁴. Die Files stehen für Forschung und den unkommerziellen Einsatz frei zur Verfügung. Jede Testsequenz besteht aus einer kurzen Filmsequenz, welche mit acht synchronisierten Basler a601fc Kameras aufgenommen wurde, die an einen PC via Firewire angeschlossen waren. Die Aufnahme erfolgte in 25 Hz mit einer Auflösung von 640x480 Pixeln (VGA). Die Testsequenzen liegen im YUV-Format vor.



Abbildung 5.6.: Samples MERL

5.4.3. Monoskope Videodateien - TU-München

Als monoskopische Testsequenzen werden die Videoclips des Lehrstuhl für Datenverarbeitung der Technische Universität München verwendet. Diese befinden sich öffentlich zugänglich auf dem FTP-Server des Fachbereichs und sind ebenfalls zur nichtkommerziellen Nutzung freigegeben⁷⁵.

Die Dateien liegen als YUV-Datei in der Größe 1920 x 1080 vor. Neben progressive encodierte Videosequenzen gibt es auch interlaced codierte Testsequenzen.

5.5. Testsoftware

Als Testsoftware für YUV-Videos wurde das Testprogramm YUVTools 3.0 der Firma Sunray Image und verwendet.

Zur Erzeugung der H264/AVC Testdatei wurde das Programm MeGUI verwendet.

Zur Erzeugung der AVCHD-Testdisc wurde das Programm multiAVCHD verwendet.

Zur Erzeugung der BD5-Testdisc wurde das Programm tsMuxerGUI verwendet.

5.6. Testhardware

5.6.1. Player

Als Testhardware wird der Standalone Bluray-Player Sony BDP S360 genutzt. Dieser spielt sowohl AVCHD-Datenträger als auch Bluray-konforme Videodaten auf herkömmlichen DVD-Datenträgern ab. Diese bezeichnet man auch als BD5 und BDD9.

⁷⁴ <ftp://ftp.merl.com/pub/avetro/mvc-testseq>

⁷⁵ ftp://ftp.ldv.e-technik.tu-muenchen.de/dist/test_sequences/

Desweiteren unterstützt der Player das Bluray Profil 1.1 und ist damit Picture in Picture (PIP) fähig. Dies wird auch als Bonusview-fähig bezeichnet.

5.6.2. Display

Zur Beurteilung der visuellen Qualität wurde das autostereoskopische LCD-Display SL2400 der Firma Tridelity verwendet (Bildschirmdiagonale 24“), welches nach dem Lentikularlinsen-Prinzip funktioniert. Zur Auswertung der monoskopischen Tests wurde ein LCD-Display vom Typ Acer AL2423W verwendet.

6. Messergebnisse

Ziel der Tests war es, eine erste grobe Einschätzung eines MVC Codecs bez. seiner Fähigkeiten zur Distribution von S3D-Material zu bekommen. Es wird daher darauf hingewiesen, dass es sich hierbei nicht um standardisierte Videoqualitätsanalysen in korrekter empirischer Methodik handelt, sondern um eine erste subjektive Annäherung an die Fähigkeiten des JMVC-Referenzcodecs zur Vorbereitung weiterführender Tests.

Auf einen weitere Gefahr der Fehlinterpretation muss ebenfalls deutlich verwiesen werden: Die benutzten Testsequenzen stellen die optimale Form von Input dar, die dieser Codec verarbeiten kann. Das L/R-Bildmaterial ist frei von jeglichen Vertikalversätzen, die L/R-Bilder weisen keine nennenswerten Schärfe-, Linsenverzeichnungs- oder chromatische Aberrations-Differenzen auf. Mit all diesen Störungseffekten müssen sich z.B. stereoskope Kameracodecs auseinandersetzen. Diese Messergebnisse sind daher in keinsten Weise dazu geeignet H.264/MVC als potenziellen Stereokamera-Codec zu bewerten, sondern beziehen sich rein auf seine Fähigkeit als Distributions-Codec von – in der Postpro maximal optimiertem – S3D-Material.

6.1. Anforderungskatalog

6.1.1. Abwärtskompatibilität

Multi-view Video ist bereits in standardmäßig durch die Einbettung eines Baseviews zu H264/AVC abwärtskompatibel. Für die Distribution von stereoskopen Filmen ist die Abwärtskompatibilität zu bestehenden Hardwareplayern äußerst wichtig, da es einerseits kaum Endgeräte gibt, die bereits in der Lage sind Multi-view Video abzuspielen. Andererseits gibt es aber bereits eine relativ große Verbreitung von H264/AVC-basierten Bluray-Playern, auf welchem Multi-view Video fehlerfrei monoskopisch wiedergeben werden muss, wenn sich MVC-basierte Bluray-Discs als neuer Standard für stereoskopische Videoinhalte etablieren soll.

Weiterhin ist es auch die Kompatibilität von Multi-view Video zu regulärer Wiedergabesoftware am PC wichtig, da heute bereits viele Videodateien am PC betrachtet und bearbeitet werden.

Das Kriterium der Abwärtskompatibilität ist gegeben, wenn bei der Wiedergabe keine Artefakte oder andere Bildfehler auftreten und das Video vom Player daher wie ein reguläres H264/AVC-Video behandelt wird.

6.1.2. Effizienz der Interview Prediktion

Um die gesteigerten Datenmengen, die bei Multi-view Video auftreten, besser zu reduzieren, werden neben der Codierung der zeitlichen Abhängigkeiten innerhalb eines Views auch die Abhängigkeiten zwischen mehreren Views eines definierten Zeitpunkts in die Datenreduktion einbezogen. Die Effizienz der Interview Prediktion wird durch einen Vergleich der PSNR-Werte von Testsequenzen mit eingeschalteter Interview

Prediktion und mit deaktivierter Interview Prediktion gemessen. PSNR bedeutet Peak Signal to Noise Ratio und zeigt die mathematische Abweichung zweier Videosignale über alle Pixelpositionen als logarithmischen Wert an. Da PSNR ein rein mathematischer Wert ist, wird in einem weiteren Plausibilitätstest die Bildqualität der codierten Testsequenzen aus einem anderen Standpunkt betrachtet: Ein positiver Effekt der Interview Prediktion liegt vor, wenn bei gleicher Videodatenrate die Testsequenz mit Interview Prediktion einen höheren PSNR-Wert aufweist als die Testsequenz ohne Interview Prediktion. Im umgekehrten Fall sollte bei gleichem PSNR-Wert eine geringere Videodatenrate bei der Testsequenz mit Interview Prediktion auftreten als bei der Testsequenzen ohne Interview Prediktion. Zusätzlich wurde bei jedem Testdurchlauf noch der SAD-Wert (sum of absolute differences) zwischen original und Codierungsergebnis beobachtet.

6.1.3. Angewendete Prediktion Struktur

Multi-view Video Coding ist sehr flexibel in der Art, wie die Interview Prediktion zwischen verschiedenen Views funktioniert. Einerseits kann nur zwischen Keyframes referenziert werden oder alternativ auch zwischen allen Arten von Frames. Damit steigen allerdings auch die Anforderungen an den Decoder.

6.1.4. Bildqualität

Da bei der Codierung mit Interview Prediktion Keyframes durch abhängige intercodierte Frames wie B-Frames ersetzt werden, kann sich theoretisch die Bildqualität dadurch verschlechtern, dass sich Bildfehler kaskadieren. Welchen Einfluss Interview Prediktion auf die Bildqualität hat, soll in diesem PSNR-Testdurchlauf ermittelt werden. Desweiteren fand noch eine rein subjektive Bildbewertung des codierten Materials an einem 24-Zoll autostereoskopischen Display (SL2400 der Firma Tridality) durch den Autor statt.

6.2. Durchführung

6.2.1. Abwärtskompatibilität

Als Testdateien für Durchführung der Test zur Abwärtskompatibilität wurden die Ballettestsequenz der Microsoft Research 3D Clips benutzt. Diese liegen in der Auflösung von 1024 Pixel mal 768 Pixel in acht separaten Views vor. Da diese Auflösung keine der auf Bluray-Disc spezifizierten Auflösung entspricht, wurden diese mit Virtualdub und Avisynth in die kompatible Auflösung von 1280 Pixel mal 720 Pixel progressiv bei 4:2:0 Chromasubsampling transponiert.

Dazu wurden folgende Avisynth-Skripte für View 0 und View 1 benutzt.

Avisynthscript für View 0

```
AVIFileSource("F:\testing\ballet0.avi")  
+AVIFileSource("F:\testing\ballet2.avi")
```



```
+AVIFileSource("F:\testing\ballet4.avi")
+AVIFileSource("F:\testing\ballet6.avi")
crop(0,24,0,-24)
AddBorders(128,0,128,0)
AssumeFPS(23.976)
ConvertToYV12()
```

```
Avisynthscript für View 1
AVIFileSource("F:\testing\ballet1.avi")
+AVIFileSource("F:\testing\ballet3.avi")
+AVIFileSource("F:\testing\ballet5.avi")
+AVIFileSource("F:\testing\ballet7.avi")
crop(0,24,0,-24)
AddBorders(128,0,128,0)
AssumeFPS(23.976)
ConvertToYV12()
```

Um die Laufzeit der Videodatei zu verlängern, wurden mehrere Videosequenzen aneinandergereiht und wiederholt. Das Video des linken Views enthält die Videobilder der Kamera 0,2,4,6 und das Video des rechten Views enthält Videobilder der Kamera 1, 3, 5, 7. So wurde sichergestellt, dass das die Testsequenz an Länge gewinnt, aber immer noch zwei benachbarte Kamerabilder in beiden Videostreams parallel codiert sind. Dies ist für die Interview Prediktion von großer Bedeutung.

Damit die Videodateien eine Bluray-Disc konforme Auflösung erhalten, wurde mit dem Befehl Crop, vertikal jeweils Oben und Unten 24 Pixel entfernt und anschließend mit dem Befehl AddBorders horizontal links und rechts jeweils 128 schwarze Pixel hinzugefügt. Anschliessend wurde die Framerate auf die bei den meisten Bluray-Disc vorherrschende Bildwiederholrate von 23.976 Hertz festgelegt und das Chroma-subsampling auf 4:2:0 festgelegt.

Anschließend wurde beide Testsequenzen mit dem Programm H264AVCEncoder-LibTestStatic, welches Bestandteil des JMV-Codecs ist, mit dem durchschnittlichen Quantisierungsparameter von QP= 19, einer GOP von 8 Videoframes und eingeschalteter Interview Prediktion nacheinander encodiert. Im Anschluss wurden beide einzelnen MVC-codierten Videostreams mit dem Programm MVCBitStreamAssembler zu einem gemeinsamen MVC-Videostream zusammengefügt. Um diesen MVC-codierten Videostream in einem Hardwareplayern abzuspielen, muss dieser erst in einen Mpeg Transportstream verpackt werden. Dies erfolgt mit dem Programm TSRemuxer, welches neben dem MPEG Transportstream auch die korrekte Ordnerstruktur einer Bluray-Disc anlegt. Zur Erstellung einer AVCHD-Testdisc wurde das Programm multiAVCHD benutzt. Die so erzeugten Dateien und Ordner werden schließlich jeweils auf eine DVD-R gebrannt.

Um die Kompatibilität mit Mediaplayern wie VideolanClient (VLC) oder Media Player Classic Home Cinema (MPC HC) zu testen, wurde der zuvor erstellte MVC-Videoströme in einen Matroska-Videocontainer verpackt. Der MKV-Container wurde mit dem Programm Mkvmerge erstellt.

6.2.2. Effizienz der Interview Prediktion

Um die Effizienz der Interview Prediktion zu messen, wurden die Balletszene der Microsoft Research 3D Testsequenzen und die Ballroomsequenz mit verschiedenen Quantisierungsgrößen codiert. In einer Testreihe war die Interview Prediktion aktiviert und in einer Testreihe deaktiviert. Als Quantisierungsparameter QP wurden die Werte 5, 10, 15, 20, 25, 30, 35 und 40 benutzt. Die GOP-Länge war auf 8 Frames beschränkt.

Nach der Codierung mit dem Encoder H264AVCEncoderLibTestStatic wurden jeweils beide encodierten Views eines Encodierdurchgangs in eine MVC-Video datei zusammengeführt. Dazu wurde das Programm MVCBitStreamAssemblerStatic verwendet. Die ausgegebene Datei anschließend durch H264AVCDecoderLibTestStatic wiederum in die einzelnen Views aufgeteilt und als YUV-Datei gespeichert⁷⁶. Danach wurde mit dem Programm PSNRStatic die PSNR-Werte einzelner Views bestimmt und in die Übersicht eingetragen.

Ballet- und Ballroomsequenz

Die PSNR-Werte der encodierten Videodaten der Ballettsequenz und Ballroomsequenz unterscheiden sich in Bezug auf den Quantisierungsparameter kaum. Bei geringen Quantisierungsparametern zwischen QP=5 und QP=20, gibt es keine Unterschiede. Erst ab einem Quantisierungsparameter von QP=30 treten Unterschiede im PSNR-Wert auf. Wie in der nachfolgenden Tabelle zu sehen ist, ist die Bitrate, die der Encoder bei Verwendung von Interview Prediktion verwendet geringer als bei Simulcast. Bei geringeren Bitraten, die ab einem Quantisierungsparameter von QP = 15 auftreten, tritt dies besonders stark auf. Dies bedeutet, dass bei rechnerisch gleichbleibender Bildqualität die Bitrate bei Verwendung von Interview Prediktion gesenkt werden kann verglichen mit Simulcast.

6.2.3. Angewendete Prediktion Struktur

Um herauszufinden, welche Prediktions-Struktur verwendet wird, wurden die Testsequenz Balletszene mit verschiedenen GOP-Längen und verschiedenen Quantisierungsparametern encodiert.

⁷⁶ Eine direkte Umwandlung der einzelnen encodierten Views in YUV-Dateien war nicht möglich, da YUV-Konverter wie YUV-Studio die MVC-codierten Videodateien nicht verarbeiten können.

	QP 05	QP 10	QP 15	QP 20	QP 25
Interview Prediktion	51,8 db	47,0 db	43,9 db	42,8 db	41,8 db
Simulkast	51,8db	47,0 db	43,9 db	42,8 db	41,9 db
Prediktionstruktur	QP 30	QP 35	QP 40		
Interview Prediktion	40,0 db	37,8 db	35,2 db		
Simulkast	40,3db	38,3 db	35,7 db		

Tabelle 6.1.: Vergleich Y-PSNR (Chrominanz)- Balletszene

	QP 05	QP 10	QP 15	QP 20	QP 25
Interview Prediktion	51,6 db	49,1 db	48,2 db	47,1 db	45,2 db
Simulkast	51,6db	47,1 db	48,2 db	47,1 db	45,5 db
Prediktionstruktur	QP 30	QP 35	QP 40		
Interview Prediktion	42,5 db	40,1	38,0 db		
Simulkast	43,1db	40,9	37,7 db		

Tabelle 6.2.: Vergleich U-PSNR (Luminanz U) - Balletszene

	QP 05	QP 10	QP 15	QP 20	QP 25
Interview Prediktion	51,7 db	49,4 db	48,6 db	47,6 db	46,3 db
Simulkast	51,7db	49,4 db	47,6 db	47,6 db	46,2 db
Prediktionstruktur	QP 30	QP 35	QP 40		
Interview Prediktion	45,0 db	43,1	41,2 db		
Simulkast	44,3db	42,4	40,8 db		

Tabelle 6.3.: Vergleich V-PSNR(Luminanz V) - Balletszene

	QP 05	QP 10	QP 15	QP 20	QP 25
Interview Prediktion	58,0 db	49,0 db	43,9 db	40,6 db	38,4 db
Simulkast	58,0 db	49,0 db	43,9 db	40,3 db	38,4 db
Prediktionstruktur	QP 30	QP 35	QP 40		
Interview Prediktion	36,0 db	33,0 db	30,0 db		
Simulkast	36,3db	33,4 db	30,6 db		

Tabelle 6.4.: Vergleich Y-PSNR (Chrominanz)- Ballroomsequenz

	QP 05	QP 10	QP 15	QP 20	QP 25
Interview Prediktion	58,6 db	49,1 db	44,2 db	43,3 db	41,3 db
Simulkast	58,6db	49,1 db	44,2 db	43,1 db	41,2 db
Prediktionstruktur	QP 30	QP 35	QP 40		
Interview Prediktion	39,5 db	37,1	36,6 db		
Simulkast	39,5db	37,5	36,7 db		

Tabelle 6.5.: Vergleich U-PSNR (Luminanz U) - Ballroomsequenz

	QP 05	QP 10	QP 15	QP 20	QP 25	
Interview Prediktion	58,67 db	49,4 db	44,6 db	42,6 db	40,3	db
Simulkast	51,7db	49,4 db	44,6 db	42,6 db	40,2	db
Prediktionstruktur	QP 30	QP 35	QP 40			
Interview Prediktion	39,0 db	36,9	36,0 db			
Simulkast	39,2db	36,4	36,4 db			

Tabelle 6.6.: Vergleich V-PSNR(Luminanz V) - Ballroomsequenz

GOP-Länge	Codiermethode	QP	PSNR	Bitrate
08 Frames	Interview Prediktion	20	47,6 db	1373 kbit/s
16 Frames	Interview Prediktion	20	42,7 db	1160 kbit/s
32 Frames	Interview Prediktion	20	42,7 db	1071 kbit/s
08 Frames	Simulkast	20	47,6 db	1534 kbit/s
16 Frames	Simulkast	20	42,8 db	1243 kbit/s
32 Frames	Simulkast	20	42,7 db	1117 kbit/s
08 Frames	Interview Prediktion	30	40 db	309 kbit/s
16 Frames	Interview Prediktion	20	40,0 db	265 kbit/s
32 Frames	Interview Prediktion	30	39,9 db	126 kbit/s
08 Frames	Simulkast	30	40,3 db	405 kbit/s
16 Frames	Simulkast	20	40,2 db	269 kbit/s
32 Frames	Simulkast	30	40,1 db	140 kbit/s

Tabelle 6.7.: Vergleich Simulkast und Interview Prediktion für GOP=8, GOP=16 und GOP=32

GOP-Länge	QP		GOP-Struktur
08 Frames	Interviewprediktion	0	I-I-B-B-B-B-B-B-I
	Interviewprediktion	1	P-P-B-B-B-B-B-B-P
16 Frames	Interviewprediktion	0	I-P-B-B-B-B-B-B-I-
			B-B-B-B-B-B-B-P
	Interviewprediktion	1	P-P-B-B-B-B-B-B-P-
32 Frames	Interviewprediktion	0	I-P-B-B-B-B-B-B-P-
			B-B-B-B-B-B-B-P-
			B-B-B-B-B-B-B-I-
			B-B-B-B-B-B-B-P
	Interviewprediktion	1	P-P-B-B-B-B-B-B-P-
			B-B-B-B-B-B-B-P-
			B-B-B-B-B-B-B-P-

			B-B-B-B-B-B-B-P-
08 Frames	Simulkast	0	I-I-B-B-B-B-B-B-I
	Simulkast	1	I-I-B-B-B-B-B-B-I
16 Frames	Interviewprediktion	0	I-P-B-B-B-B-B-B-I-
			B-B-B-B-B-B-B-P
	Interviewprediktion	1	I-P-B-B-B-B-B-B-I-
			B-B-B-B-B-B-B-P
32 Frames	Interviewprediktion	0	I-P-B-B-B-B-B-B-B-P-
			B-B-B-B-B-B-B-P-
			B-B-B-B-B-B-B-I-
			B-B-B-B-B-B-B-P
	Interviewprediktion	1	I-P-B-B-B-B-B-B-B-P-
	Interviewprediktion	1	I-P-B-B-B-B-B-B-B-P-
			B-B-B-B-B-B-B-P-
			B-B-B-B-B-B-B-I-
			B-B-B-B-B-B-B-P

Tabelle 6.8.: Vergleich GOP=8, GOP=16 und GOP=32

6.2.4. Bildqualität

Zur Bestimmung der Bildqualität wurden die encodierten Videodateien der Testsequenz Ballettszene verwendet. Im ersten Schritt wurde die Videodateien verwendet, die mit der GOP-Länge von 8 und verschiedenen Quantisierungsparameter encodiert wurden. Im zweiten Schritt wurden Clips verschiedener GOP-Längen mit konstanten Quantisierungsparameter verglichen. Als Testsoftware wurde das Programm YUV-Studio verwendet. Dieses bietet die Funktion Comparison mit welche zwei YUV-Videodateien verglichen werden können und das Ergebniss direkt visuell angezeigt wird.

Zur Evaluation der Bildqualität wurden zuerst der codierte View 0 mit der Originalvideodatei monoskopisch verglichen. Da View 0 sowohl bei eingeschalteter Interview Prediktion als auch bei Simulcast identisch ist, bietet der Vergleich eine allgemeine Beurteilung der Auswirkung des Quantisierungsparameters auf die Codierung. Im zweiten Schritt, wurde View 1 der encodierten Dateien mit aktivierter Interview Prediktion mit View 1 der Simulcast-Variante verglichen. Dies sollte aufzeigen, ob es Unterschiede in der visuellen Qualität zwischen Interviewprediktion und Simulcast gibt.

6.2.4.1. Monoskopischer Vergleich Original vs. View 0

QP5

Die Testsequenz wurde mit einem Quantisierungsparameter von $QP = 5$ codiert. Bei einer Bitrate von 42100 kbit/s hat die Testsequenz einen PSNR-Wert von 51,8 db. Visuell ist keine Artefaktbildung zu erkennen. Die Testsequenz ist visuell von der unkomprimierten Originaldatei praktisch nicht zu unterscheiden. Auch der SAD-Wert zeigt keine Unterschiede an.

Q10

Die Testsequenz wurde mit einem Quantisierungsparameter von $QP = 10$ codiert. Bei einer Bitrate von 18446 kbit/s hat die Testsequenz einen PSNR-Wert von 47,0 db. Visuell ist keine Artefaktbildung zu erkennen. Von der unkomprimierten Originaldatei ist die Testsequenz visuell praktisch nicht zu unterscheiden. Auch die Visualisierung der SAD-Werte zeigt keine Differenzen an.

Q15

Die Testsequenz wurde mit einem Quantisierungsparameter von $QP = 15$ codiert. Bei einer Bitrate von 4784 kbit/s hat die Testsequenz einen PSNR-Wert von 43,9 db. Visuell ist keine Artefaktbildung zu erkennen. Die Testsequenz ist praktisch von der unkomprimierten Originaldatei nicht zu unterscheiden. Die Visualisierung der SAD-Werte zeigt nur ganz leichte Veränderungen an.

Q20

Die Testsequenz wurde mit einem Quantisierungsparameter von $QP = 20$ codiert. Bei einer Bitrate von 1373 kbit/s hat die Testsequenz einen PSNR-Wert von 42,8 db. Visuell ist gegenüber der unkomprimierten Originaldatei eine leichte Verschlechterung des Bildeindrucks vorhanden. Das Bild wirkt leicht weichgezeichnet und etwas detailärmer. Der in der Testsequenz dargestellte Fußboden ist etwas weniger strukturiert und das in der Originaldatei leicht vorhandene Rauschen ist verschwunden. Artefaktbildung tritt nicht auf. Die Visualisierung der SAD-Werte zeigt eine begrenzte Verschlechterung gegenüber der Originaldatei an.

Q25

Die Testsequenz wurde mit einem Quantisierungsparameter von $QP = 25$ codiert. Bei einer Bitrate von 594 kbit/s hat die Testsequenz einen PSNR-Wert von 41,8 db. Visuell ist gegenüber der unkomprimierten Originaldatei eine stärkere Verschlechterung des Bildeindrucks zu erkennen. Das Bild wirkt stärker weichgezeichnet und flächig. Vor allem der Fußboden ist stark weichgezeichnet. Das Bildrauschen ist komplett verschwunden. Artefakte treten nicht auf. Die Visualisierung der SAD-Werte zeigt eine stärkere Verschlechterung gegenüber der Originaldatei an.

Q30

Die Testsequenz wurde mit einem Quantisierungsparameter von $QP = 30$ codiert. Bei einer Bitrate von 309 kbit/s hat die Testsequenz einen PSNR-Wert von 40,0 db. Visuell ist gegenüber der unkomprimierten Originaldatei eine starke Verschlechterung des Bildeindrucks vorhanden. Das Bild wirkt stark weichgezeichnet und strukturarm. Artefakt treten im Bereich des Bodens und am Körper der Tänzerin auf. Besonders an den Armen der Tänzerin kommt es zu Artefakten. Dadurch verschmelzen Hintergrund und Vordergrund leicht. Auch die Visualisierung der SAD-Werte zeigt eine stärkere Verschlechterung an den Konturen beider Figuren an.

Q35

Die Testsequenz wurde mit einem Quantisierungsparameter von $QP = 35$ codiert. Bei einer Bitrate von 137 kbit/s hat die Testsequenz einen PSNR-Wert von 37,8 db. Visuell ist gegenüber der unkomprimierten Originaldatei eine sehr starke Verschlechterung des Bildeindrucks zu erkennen. Das Bild wirkt äußerst weichgezeichnet und ohne Struktur.

Der Fußboden ist nahezu eine einfarbige Fläche mit starker Artefaktbildung. Der Körper der Tänzerin weist ebenfalls eine stark Artefaktbildung auf. Die Tänzerin lässt sich noch schlechter vom Hintergrund unterscheiden als zuvor.

Neben der Veränderung an den Konturen der Figuren, hat es auch eine Verschlechterung auf den flächigen Bildelementen, wie der Wand im Hintergrund, gegeben.

Q40

Die Testsequenz wurde mit einem Quantisierungsparameter von $QP = 40$ codiert. Bei einer Bitrate von 103 kbit/s hat die Testsequenz einen PSNR-Wert von 35,7 db. Visuell ist gegenüber der unkomprimierten Originaldatei eine extrem starke Verschlechterung des Bildeindrucks vorhanden. Das Bild weist keine Details mehr auf und besteht nahezu nur noch aus Artefakten. Die Trennung zwischen Vordergrund und Hintergrund ist nahezu nicht mehr gegeben. Vor allem im Bereich der Arme der Tänzerin tritt dies auf. Auch die Visualisierung der SAD-Werte zeigt eine extreme Verschlechterung im gesamten Bild an.

QP 05



QP 10



QP 15

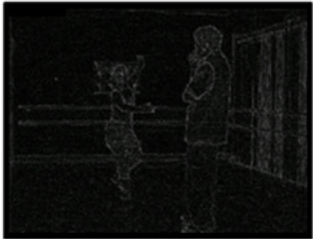


QP 20



Abbildung 6.1.: Visueller Vergleich QP 5 bis QP 20

QP 25



QP 30



QP 35



QP 40

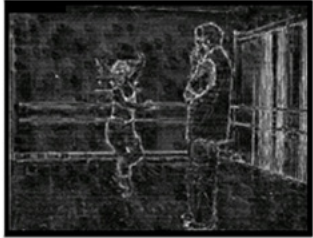


Abbildung 6.2.: Visueller Vergleich QP 20 bis QP 40

6.2.4.2. Monoskopischer Vergleich Simulcast vs. Interview Prediktion

QP5

Beide Testsequenzen wurden mit einem Quantisierungsparameter von $QP=5$ codiert. View 1 mit Interviewcodierung erreichte bei der Bitrate von 42100 kbit/s einen PSNR-Wert von 51,8 db, während View 1 in Simulcast den gleichen PSNR-Wert mit einer Bitrate von 42157 kbit/s erreichte. Zwischen beiden Videodateien ist visuell kein Unterschied zu erkennen. Auch die Visualisierung der SAD-Werte zeigt keinerlei Unterschied an.

Q10

Beide Testsequenzen wurden mit einem Quantisierungsparameter von $QP=10$ codiert. View 1 mit Interviewcodierung erreichte bei der Bitrate von 18446 kbit/s einen PSNR-Wert von 47,0 db, während View 1 in Simulcast den gleichen PSNR-Wert mit einer Bitrate von 18511 kbit/s erreichte. Zwischen beiden Videodateien ist visuell kein Unterschied zu erkennen. Auch die Visualisierung der SAD-Werte zeigt keinerlei Unterschied an.

Q15

Beide Testsequenzen wurden mit einem Quantisierungsparameter von $QP=15$ codiert. View 1 mit Interviewcodierung erreichte bei der Bitrate von 4784 kbit/s einen PSNR-Wert von 43,9 db, während View 1 in Simulcast den gleichen PSNR-Wert mit einer Bitrate von 4857 kbit/s erreichte. Zwischen beiden Videodateien ist visuell kein Unterschied zu erkennen. Die Visualisierung der SAD-Werte zeigt nur äußerst geringe Veränderungen zwischen beiden Views an.

Q20

Beide Testsequenzen wurden mit einem Quantisierungsparameter von $QP=20$ codiert. View 1 mit Interviewcodierung erreichte bei der Bitrate von 1373 kbit/s einen PSNR-Wert von 42,8 db, während View 1 in Simulcast den gleichen PSNR-Wert mit einer Bitrate von 1534 kbit/s erreichte. Zwischen beiden Videodateien ist visuell kein Unterschied zu erkennen. Die Visualisierung der SAD-Werte zeigt an den Konturen der Figuren Veränderungen zwischen beiden Views an.

Q25

Beide Testsequenzen wurden mit einem Quantisierungsparameter von $QP=25$ codiert. View 1 mit Interviewcodierung erreichte bei der Bitrate von 594 kbit/s einen PSNR-Wert von 41,8 db, während View 1 in Simulcast bei einer Bitrate von 723 kbit/s einen PSNR-Wert von 41,9 db erreichte. Zwischen beiden Videodateien ist visuell nur ein äußerst geringer Unterschied festzustellen. Die mit aktivierter Interviewprediktion-korrektur codierte Testsequenz scheint schärfer als die Testsequenz mit deaktivierter Interviewprediktion. Die Visualisierung der SAD-Werte deutet stärkere Veränderungen zwischen beiden Views an.

Q30

Beide Testsequenzen wurden mit einem Quantisierungsparameter von $QP=30$ codiert. View 1 mit Interviewcodierung erreichte bei der Bitrate von 309 kbit/s einen PSNR-Wert von 40,0 db, während View 1 in Simulcast bei einer Bitrate von 405 kbit/s einen PSNR-Wert von 40,3 db erreichte. Zwischen beiden Videodateien ist ein geringer Unterschied feststellbar. Die mit aktivierter Interviewprediktionkorrektur codierte Testsequenz erscheint schärfer als die Testsequenz mit deaktivierter Interviewprediktion. Allerdings besitzt Erstere auch mehr Artefakte, die den Bildeindruck stören. Die Visualisierung der SAD-Werte zeigte stärkere Veränderungen in allen Bildbereichen zwischen beiden Testsequenzen an.

Q35

Beide Testsequenzen wurden mit einem Quantisierungsparameter von $QP=35$ codiert. View 1 mit Interviewcodierung erreichte bei der Bitrate von 137 kbit/s einen PSNR-Wert von 37,8 db, während View 1 in Simulcast bei einer Bitrate von 243 kbit/s einen PSNR-Wert von 38,3 db erreichte. Zwischen beiden Videodateien ist ein geringer visueller Unterschied feststellbar. Die mit aktivierter Interviewprediktionkorrektur codierte Testsequenz erscheint schärfer und detailreicher als die Testsequenz mit deaktivierter Interviewprediktion. Allerdings besitzt Ersteres auch mehr Artefakte, die den Bildeindruck stören. Die Visualisierung der SAD-Werte zeigt starke Veränderungen zwischen beiden Testsequenzen an, vor allem an den Konturen der Personen und in den flächigen Bildbereichen.

Q40

Beide Testsequenzen wurden mit einem Quantisierungsparameter von $QP=40$ codiert. View 1 mit Interviewcodierung erreichte bei der Bitrate von 103 kbit/s einen PSNR-Wert von 38,0 db, während View 1 in Simulcast bei einer Bitrate von 167 kbit/s einen PSNR-Wert von 37,0 db erreichte. Zwischen beiden Videodateien ist ein geringer visueller Unterschied feststellbar. Die mit aktivierter Interviewprediktionkorrektur codierte Testsequenz scheint schärfer und detailreicher als die Testsequenz mit deaktivierter Interviewprediktion. Allerdings besitzt Ersteres auch mehr Artefakte, die den Bildeindruck stören. Die Visualisierung der SAD-Werte zeigt starke Veränderungen zwischen beiden Testsequenzen in praktisch allen Bildbereichen auf.

6.2.4.3. Stereovergleich Simulcast vs. Interview Prediction

Zum Vergleich des Stereoeindrucks wurde das Display FSL2400 der Firma Tridality benutzt. Der Abstand zum Display betrug ca 80cm. Nach einer kurzen Eingewöhnungsphase wurden die Testsequenzen nacheinander verglichen.

Q05

Bei der Betrachtung der encodierten Testsequenz konnte keine Veränderung zum Original wahrgenommen werden. Der stereoskopische Bildeindruck wurde durch die Encodierung nicht verändert.

Q10

Bei der Betrachtung der encodierten Testsequenz konnte keine Veränderung zum Original wahrgenommen werden. Der stereoskopische Bildeindruck wurde durch die Encodierung nicht verändert.

Q15

Bei der Betrachtung der encodierten Testsequenz konnte keine Veränderung zum Original wahrgenommen werden. Der stereoskopische Bildeindruck wurde durch die Encodierung nicht verändert.

Q20

Bei der Betrachtung der encodierten Testsequenz konnte - abgesehen von einer leichten Verschlechterung der Bildqualität (Weichzeichnen) - keine Veränderung des stereoskopischen Tiefeneindrucks wahrgenommen werden.

Q25

Bei der Betrachtung der encodierten Testsequenz konnte - abgesehen von einer Verschlechterung der Bildqualität durch Artefaktbildung - keine Veränderung des stereoskopischen Tiefeneindrucks wahrgenommen werden.

Q30

Bei der Betrachtung der encodierten Testsequenz konnte - abgesehen von einer starken Verschlechterung der Bildqualität durch Artefaktbildung und fehlenden Bilddetails - keine Veränderung des stereoskopischen Tiefeneindrucks wahrgenommen werden.

Q35

Bei der Betrachtung der encodierten Testsequenz konnte - neben einer weiteren Verschlechterung der Bildqualität durch Artefaktbildung und fehlenden Bilddetails - eine geringe Veränderung des stereoskopischen Tiefeneindrucks wahrgenommen werden. Diese trat gleichermaßen bei der Testsequenz mit eingeschalteter Interviewprediktion wie auch bei ausgeschalteter Interframeprediktion auf.

Q40

Bei der Betrachtung der encodierten Testsequenz konnte - neben der sehr schlechten Bildqualität durch die äußerst starke Artefaktbildung und nahezu komplett fehlenden Bilddetails - eine weitere Veränderung des stereoskopischen Tiefeneindrucks wahrgenommen werden. Dieser Effekt trat ebenso wieder bei beiden Testsequenzen auf.

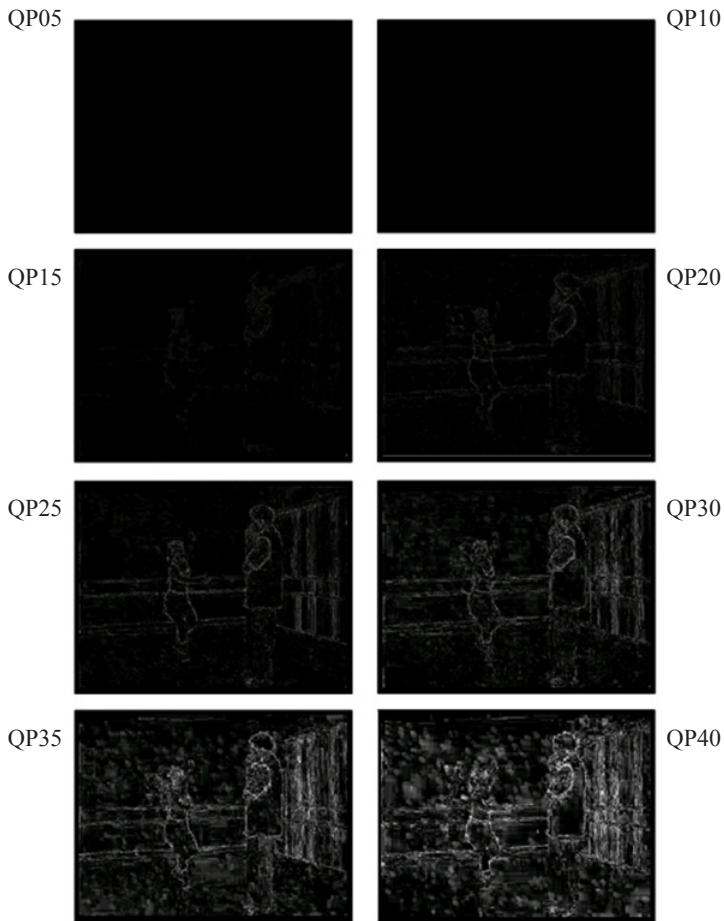


Abbildung 6.3.: Visueller Vergleich QP 20 bis QP 40

6.3. Abwärtskompatibilität

6.3.1. Hardware

Die beiden Testdiscs mit dem MVC-codierten Videoströmen wurde vom Hardware-player Sony BDP S360 problemlos monoskopisch wiedergegeben. Der Player spielte sowohl die AVCHD-Testdisc als auch die BD5-Testdisc ab. Es kam zu keiner Artefaktbildung oder Überlagerung von Inhalten des zweiten Views. Dieser wurde komplett vom Player ignoriert.

6.3.2. Software

Softwarebasierte Bluray-Player wie PowerDVD oder WinDVD spielten dagegen beide Testdiscs nicht ab. WinDVD konnte die Testdiscs nicht einlesen und brach mit der Fehlermeldung, dass das Medium nicht unterstützt werden würde, ab. PowerDVD konnte die Testdisc zwar lesen und erkannte die MVC-codierte Videosequenz als H264/AVC-Videoclip, stellte es aber nur als schwarzes Standbild dar.

Im Gegensatz zu softwarebasierten Bluray-Playern spielten Allround-Mediaplayer wie MPC HC und VLC den MVC-codierten Videostrom problemlos im Matroska-Videocontainer ab. Beide Player erkannten die Videodatei als monoskopische AVC-Videodatei und ignorierten den zweiten View. Auch hier kam es bei der Wiedergabe keine Artefakte wahrzunehmen.

6.4. Fazit

Getestet wurde der JMVC-Referenzcodec in der Version 7.0. Als Input wurde störungsfreies, von geometrischen Fehlern und Verzeichnungen befreites, unkomprimiertes Stereo-HD-Videomaterial im Y, Cb, Cr-Format verwendet. Folglich beziehen sich alle Testaussagen auf die Verwendung des MVC-Codecs als S3D-Distributionscodec für Stereo-HD-Material. Eine Aussage über MVC als Kameracodec kann nicht getroffen werden.

Betrachtet wurde zunächst der PSNR-Wert der Codierungsergebnisse bei unterschiedlichen Zieldatenraten. Desweiteren fand eine subjektive Bildbewertung der Codierungsergebnisse durch den Autor sowohl monoskopisch als auch an einem 24-Zoll Autostereodisplay statt. Es wird daher darauf hingewiesen, dass es sich nicht um standardisierte Videoqualitätsanalysen in korrekter empirischer Methodik handelt, sondern um eine erste subjektive Annäherung an die Fähigkeiten des JMVC-Referenzcodecs zur Vorbereitung weiterführender Tests.

Unter diesen Randbedingungen kann als Fazit festgehalten werden:

- Für die auf der Bluray-Disc verwendeten Zieldatenraten ist bei optimalem Input-Material (z.B. aus einer S3D-Postpro-Pipeline) - von keinerlei Qualitätsbeeinträchtigungen sowohl der allgemeinen Bildanmutung wie auch des stereoskopischen Tiefeneffekts durch die MVC-Codierung auszugehen.
- Selbst bei niedrigeren Zieldatenraten bis zu 1Mbit/s (wie sie etwa beim IP-Streaming vorkommen) konnte z.B. an einem 24"-Display keine subjektive Qualitätsverschlechterung wahrgenommen werden. Der PSNR-Wert fiel nicht unter 40db. Der Codec-Input war allerdings von hervorragender Qualität.
- Bei Zieldatenraten unter 1Mbit/s verschlechterte sich die Bildanmutung auch bei diesem guten Material deutlich durch Blockartefakte (PSNR-Werte um 36db), der subjektive stereoskopische Tiefeneindruck blieb aber noch sehr lange robust aufrecht.

Literaturverzeichnis

- 3D Vision Blog [What to Choose: IMAX 3D versus RealD versus Dolby 3D for 3D Movies?]: What to Choose: IMAX3D versus RealD versus Dolby 3D for 3D Movies? 2009 URL: <http://3dvision-blog.com/what-to-choose-imax-3d-versus-reald-versus-dolby-3d-for-3d-movies/> - Zugriff am 05.02.2010
- Bal, Can [Three-Dimensional Video Coding on Mobile]: Three-Dimensional Video Coding on Mobile Platforms. Jur. Diss., 2009
- Blu-ray Disc Association [Blu-ray Disc Association Provides Updates on 3D Technology]: 2009 URL: <http://us.blu-raydisc.com/> - Zugriff am 05.01.2010
- Eisert, Peter et al. [Free View Point Video]: Free View Point Video - der nächste Schritt in der Unterhaltungselektronik. 2006 URL: <http://iphome.hhi.de/eisert/papers/fktg2006.pdf> - Zugriff am 08.02.2010
- Gazetas, Aristides [An Introduction to World Cinema]: An Introduction to World Cinema. 2008, ISBN 0786439076
- Hart/Steve [A new dimension for JDSU]: A new dimension for JDSU. 2010 URL: <http://www.pressdemocrat.com/article/20100107/BUSINESS/100109653/1036?Title=A-new-dimension-for-JDSU> - Zugriff am 09.01.2010
- Hotton, Nikolaus [Digitale AV-Technik]: Digitale AV-Technik WS 09/10. 2009 <http://www.stereo3d.com> [<http://www.stereo3d.com> 2005 - Sensio format]:
- Janssen, Jan-Keno [3D-Standard für DVD]: 3D-Standard für DVD. 2009 URL: <http://www.heise.de/newsticker/meldung/3D-Standard-fuer-DVD-203047.html> - Zugriff am 04.01.2010
- Janssen, Jan-Keno [c't 15/2009 - Räumlich am Rechner]: c't 15/2009 - Räumlich am Rechner. [2009], 84-89
- Joint Video Team [MVC Software Manual]: MVC Software Manual. 2009
- Judge, Arthur W. [Stereoscopic Photography]: Stereoscopic Photography. 2007, ISBN 1406771724
- Kay, Jeremie [Chicken Little to get the 3-D treatment Down Under]: Chicken Little to get the 3-D treatment Down Under. 2005 URL: <http://www.screen-daily.com/chicken-little-to-get-the-3-d-treatment-down-under/> 4025569.article - Zugriff am 23.12.2009
- Lipton, Lenny [Foundations of the Stereoscopic Cinema]: Foundations of the Stereoscopic Cinema. 1982, ISBN 0442247249
- Maddox, Garry [Specs sells: Why 3D Movies have become the Real Thing]: Specs sells: Why 3D Movies have become the Real Thing. 2009 URL: <http://www.watoday.com.au/digital-life/digital-life-news/specs-sells-why-3d-movies-have-become-the-real-thing-20090914-fmpd.html> - Zugriff am 05.01.2010
- Marien, Mary Warner [Photography]: Photography. 2002, ISBN 0-13-183297-2
- Martinez-Ibañez, Jose M. [MPEG 3DAV AhG Activities Report]: MPEG 3DAV AhG Activities Report. 2003 URL: <http://www.iti.gr/SCHEMA/preview>

- html?file_id=106 - Zugriff am 05.02.2010
- Matt Cowan [RealD 3D Theatrical System]: RealD 3D Theatrical System. 2007 URL: http://www.edcf.net/edcf_docs/real-d.pdf - Zugriff am 16.02.2010
 - Mendiburu, Bernard [3D Movie Making]: 3D Movie Making: Stereoscopic Digital Cinema from Script to Screen. 2009
 - Merkle, Philipp et al. [Efficient Prediction Structures for Multi-view Video Coding]: Efficient Prediction Structures for Multi-view Video Coding. 2006 URL: http://ip.hhi.de/imagecom_G1/assets/pdfs/ieee07_Prediction_MVC.pdf - Zugriff am 20.02.2010
 - Ozaktas, Haldun M./Onural, Levent [Three-Dimensional Television]: Three-Dimensional Television. Jur. Diss., 2008
 - Pereira, Fernando et al. [Distributed Coding for Video Services]: Distributed Coding for Video Services. 2006 URL: <http://www.discoverdvc.org/deliverables/Discover-D7.pdf> - Zugriff am 19.12.2009
 - Richardson, Iain E. [H.264 and MPEG-4 Video Compression]: H.264 and MPEG4 Video Compression - Video Coding for Next-Generation Multimedia. 2008
 - Röder, Oliver [Grundlagen der Stereoskopie]: Grundlagen der Stereoskopie Analyse der Aufnahme und Projektion von 3-D Bildern. 2007, ISBN 9783836411820
 - Scott, Allen John [On Hollywood]: On Hollywood - The Place, The Industry. 2005, ISBN 9780691116839
 - Taylor, Jim [DVD Demystified]: DVD Demystified. 2001, ISBN 0071350268
 - Umble, Eden Ashley [Making It Real: The Future of Stereoscopic 3D Film Technology]: Making It Real: The Future of Stereoscopic 3D Film Technology. 2008 URL: <http://www.siggraph.org/publications/newsleter/volume-40-number-1/makingitreal> - Zugriff am 10.01.2010
 - Weise/Elizabeth [What happens to those 3-D glasses after 'Avatar'?]: What happens to those 3-D glasses after 'Avatar'? 2010 URL: http://www.usatoday.com/tech/science/environment/2010-02-02-Avatar02_ST_N.htm - Zugriff am 12.02.2010
 - Wikipedia - Pullrich-Effekt [Pullrich-Effekt]: www.Stereoforum.org [Anaglyphes 3D in zwei Streams konvertieren]: Anaglyphes 3D in zwei Streams konvertieren. 2010 URL: <http://www.stereoform.org/viewtopic.php?f=13&t=919>
 - Ying Chen et al. [The Emerging MVC Standard]: The Emerging MVC Standard for 3D Video Services. 2008 URL: <http://www.hindawi.com/journals/asp/2009/786015.abs.html> - Zugriff am 23.01.2010
 - Zone, Ray [Stereoscopic cinema and the origins]: Stereoscopic cinema and the origins of 3-D film, 1838-1952. Lexington, Ky: The University Press of Kentucky, 2007 URL: <http://www.loc.gov/catdir/toc/ecip0714/2007012691.html>, ISBN 9780813124612

Teil 7

S3D-Grammatik: Analysen, Einordnungen und Definitionsansätze zur Bildsprache des stereoskopischen Films

Paul Walter

Abbildungsverzeichnis

Abbildung 1: Vergleich der Umwandlung von Film in ein Fernsehbild.....	409
Abbildung 2: Figuren von Marvin L. Misky und Seymour A. Pappert	414
Abbildung 3: Paradigmatische und syntagmatische Konnotationsprozesse	421
Abbildung 4: Schaubild kinematografischer Codes	422
Abbildung 5: Einordnung der Stereografie als neuer kinematografischer Code ..	424
Abbildung 6: Die drei Kompositionsebenen nach Monaco	427
Abbildung 7: Szenen aus Inglourious Basterds	431
Abbildung 8: Tiefenscript von Brian Gardner	435

1 Einleitung

Mit Beginn dieses Jahrzehnts muss konstatiert werden, dass Filmproduzenten, Verleiher und Kinobetreiber die Scheu vor dem sog. „Digitalen Kino-Standard“ der DCI (Digital Cinema Initiative) abgelegt haben.

Im Gegenteil: um dem Problem der fallenden Kinobesuchszahlen entgegen zu wirken, soll „3D-Fähigkeit“ das Schlagwort sein um Zuschauer (wieder einmal) vermehrt zurück in die Kinos zu locken. Ein aktueller Trend im digitalen Kino geht daher deutlich zum 3D-Stereofilm, was sich auch in der Öffentlichkeitsarbeit der Produzenten und Verleiher deutlich zeigt. Zum Beispiel wurde der stereoskopische 3D-Animationsfilm *OBEN* (Pixar, 2009) zum Eröffnungsfilm der Filmfestspiele in Cannes gekürt. Ähnliches geschah im September 2009, als auf dem Filmfestival in Venedig erstmals Preise für die besten 3D-Stereofilme vergeben wurden. Im Frühjahr 2010 waren weltweit etwa 60 3D-Stereofilme in Produktion und ihre Zahl steigt weiter kontinuierlich an. Die Verkaufszahlen von 3D-fähigen Displays steigen ebenfalls nachhaltig. Der 3D-Stereofilm steht also vor einer Reaktivierung, aber die Gefahr besteht natürlich weiterhin, dass er nach einem Hype (wie so oft) in der Versenkung verschwindet¹.

Eine Studie der „Media Perspektiven“ im Auftrag der Hochschule für Film und Fernsehen in Potsdam zeigt, dass der Trend sich aber auch fortsetzen könnte. Insgesamt wurden 1002 Personen im Alter von 16 und 64 Jahren telefonisch befragt. Insgesamt gaben 44,4 Prozent der Befragten an, das sie schon einen 3D-Film gesehen haben. Als erster Berührungspunkt wird häufig das IMAX-Kino angegeben. Die Befragten zeigten durchweg Interesse an der neuen Technik, 53,9 Prozent der Männer und 51,7 Prozent der Frauen haben sogar großes Interesse an 3D. Das noch notwendige Übel

1 Vgl. Marks Chip 2009, 2009.

der 3D-Brille stellt nicht unbedingt eine Barriere dar. 54,4 Prozent der Befragten gaben an, das sie die Brille kaum stört.

Dass 3D-Projektionen mit höheren Eintrittsgeldern verbunden sind, wird von 46,1 Prozent der Befragten als gerechtfertigt empfunden. Bei Jugendlichen bis 19 Jahren, lag die Zahl sogar bei 70 Prozent, was darauf deutet in dieser Gruppe die Bereitschaft für einen stereoskopischen Kinofilm mehr zu bezahlen höher ist². Diese Akzeptanz war dem Kino in den letzten Jahren eher abhanden gekommen.

Aufgrund dieser Entwicklung soll in diesem Artikel - jenseits von systemtechnischen und ökonomischen Aspekten - zunächst folgende Frage bearbeitet werden: Wird mit Stereofilmen ein inhaltlicher Mehrwert für den Zuschauer generiert und welcher ist das?

Falls diese Frage, die sich nur auf die erzählerische bzw. kommunikative Wirkung eines Films bezieht, ansatzweise positiv beantwortet werden könnte, wäre davon auszugehen, dass der 3D-Stereofilm mehr ist als nur ein „digitales Strohfeuer“ sondern tatsächlich eine weiterer Schritt hin zur wirkungsvollen Kommunikation der Produzenten mit dem Zuschauer.

Für die Herstellung solcher Filme ist weithin mehr nötig als eine gute Story und dem Aspekt dass es sich zum Schluss um eine Stereoprojektion handelt. Sie hat den Auftrag die verfügbaren filmischen Stilmittel (Codes) in handwerklich und dramaturgisch sinnvollen Regeln einzusetzen und betritt damit spannendes Neuland. Wie sieht der tatsächliche dramaturgische Nutzen für die Filmschaffenden aus? Dieser Artikel versucht daher auch im Spannungsfeld von Technik, Wahrnehmungspsychologie und Bildsprache bzw. kinematografischer Grammatik erste Antworten bezüglich Stereo-Kreativität und sinnvollem Stereo-Handwerk zu finden.

Ein Blick in die Vergangenheit zeigt eins nämlich deutlich: dass der Zuschauer handwerklich schlecht produzierte Stereofilme – losgelöst vom Plot - unmittelbar und heftig ablehnt.

2 Vgl. Jockenhövel 3D Kino im Urteil des Publikums, 2009.

2. Historische Entwicklungszyklen im Kino

2.1 Tonfilm

Fortschritt heißt Bewegung und diese geht meist nicht ohne Abschied einher, deswegen kann man bei allen neuen Erungenschaften des Films immer auch einen Verlust feststellen. Wenn der Stummfilm als Beispiel aufgeführt wird, hat dieser durch den Einfluss den Tons (und aufgrund der damit neu aufkommenden Montagetechniken) den künstlerischen Ausdruck des Schauspiels verändert. Es kam in Folge zu mimischen Ausdrucksverlusten³. Das überbetonte Agieren wich einer subtileren Mimik und Gestik. Der Ton verbindet das Gesehene und sorgte dadurch beim Zuschauer für ein kompakteres Erlebnis und erleichtert die Erzählung. "Wenn wörtliche Rede eine erklärende, sinnvermittelnde Funktion zugewiesen bekommt, werden Mimik und Gestik entlastet. Doch zu Beginn des Tonfilms war es häufig sogar so, dass sich die Schauspieler erst an den Ton selbst gewöhnen mussten. Sie waren es nicht gewohnt ihr Schauspiel mit einem Ton zu verbinden und es wirkte Vieles zunächst sehr gestellt, weil sie darauf bedacht waren ihren Text verständlich zu "sprechen"⁴. Hinter der damaligen Erfindung des Tonfilms standen natürlich auch ökonomische Gründe und seine Einführung ging besonders in den USA mit einem dramatischen Wandel des Studiosystems einher. Die neue technische Entwicklung hatte durch die notwendigen Umrüstungen auch einen hohen Produktionsaufwand zu Folge. Zusätzlich gab es Probleme mit der Aufnahme von Ton direkt am Drehort bzw. Studio. Die Lautstärke der mechanischen Kamera sorgte für Probleme bei der Tonaufnahme, doch diese Probleme wurden im Laufe der Zeit gelöst, weil erkannt wurde, welche gestalterische Potentiale im Tonfilm steckten. Beispielhaft erwähnt seinen weiche Schnitte mitten im Dialog, die auditive Fortführung von Aktionen außerhalb des sichtbaren Bildbereichs und natürlich die Erweiterung durch Musik, Klang und Geräusche als dramaturgisches Element⁵.

Der Tonfilm wurde letztendlich ein wichtiges Element im Film und kein Effekt. Heute sind es z.T. ganze Filmgenres, die durch die musikalische Untermalung überhaupt erst zum Leben erwachen.

2.2 Farbfilm

Der Farbfilm ist für uns heute ebenfalls zur Normalität geworden, doch auch er hat einen langen und hindernisreichen Entwicklungsweg durchgemacht bis er akzeptiert wurde. Erste Entwicklungen wurden in den 1910er und 1920er Jahren gemacht, je-

3 Vgl. Schnell Medienästetik - Zu Geschichte und Theorie audiovisueller Wahrnehmungsformen, 2000, S. 94.

4 D. Bordwell Visual Style in Cinema - Vier Kapitel Filmgeschichte, 2006, S. 113.

5 Vgl. David Bordwell Film Art: An Introduction, 2003, S. 97.

doch erst in den 1930er Jahren durch die Entwicklung von Technicolor⁶ hat sich ein Verfahren durchgesetzt, dass die bewegte Farbfilmproduktion revolutionierte. Farbe konnte nun zum ersten Mal relativ störungsfrei als dramaturgisches Mittel eingesetzt werden⁷. Zwar war das Farbmateriale oft noch nicht lichtempfindlich genug und es entstanden Lichtschwankungen und Brüche in der Farbkontinuität, die Wirkung auf den Zuschauer war aber meist überwältigend.

Trotz der offensichtlichen Wirkung von Farbe gab es jedoch auch kritische Stimmen die besagten, dass Farbe eine „Verflachung der Bildes“ zur Folge haben wird. Dabei wurde speziell die Realitätsnähe der Bilder als Hauptgrund genannt, die von den Filmemachern und Zuschauern angeblich nicht gewollt wurde. Hinzu kamen noch ökonomische Gründe, da die Produktion in Farbe die Kosten in die Höhe trieb. Zu diesen Problemen kamen dann noch ungeklärte Gestaltungsfragen hinzu, da die Arbeit mit Farben noch Neuland war⁸. Dies alles sorgte dafür, dass der Farbfilm nur sehr selten und wenn dann äußerst selektiv für bestimmte Szenen oder Genres genutzt wurde. Die Farbfilmtechnik sorgte allerdings auch dafür, dass sich neue Möglichkeiten in der Filmgestaltung auftaten. DAVID BORDWELL beschreibt in seinen Untersuchungen, dass der „technischen Bewältigung des Farbproblems auf einer anderen Ebene neue Formen der erzählerischen und visuellen Gestaltung von Filmen gegenüberstanden“⁹. Sie sorgten für eine Erweiterung filmischer Erzählweisen und bedingten neue Eigenschaften der Montage¹⁰.

2.3 Breitwandfilm

Da die Entwicklung des sog. Mainstream-Kinos in erster Linie an ökonomischen Faktoren ausgerichtet ist denn an ästhetischen Aspekten, hat dies meist zur Folge, dass es immer eine gewisse Zeit dauert bis sich eine technische Veränderung dauerhaft etablieren kann. Beim „Breitwand“-Filmformat war dies nicht anders. Es war zwar schon ziemlich früh in der Geschichte des Kinos deutlich, dass die Breite der Projektion nicht mit dem Gesichtsfeld des Menschen übereinstimmt - hauptsächlich bedingt durch das 35mm Bildformat und die vier Perforationslöcher, die auf dem Bildstreifen ausgerichtet sind. Dabei gab es schon früh Erfindungen, wie das „Anamorphoscope“-Verfahren¹¹ von HENRI CHRÉTIEN, der die Aufnahme und Wiedergabe von Breitwandbildern auf 35mm Film erlaubte. Anfang der 1950er Jahre griff die Filmindustrie abermals

6 Verfahren für die Aufnahme und Wiedergabe von Farbfilmen, es wurden zwischenzeitlich insgesamt fünf Verfahren entwickelt, richtig durchgesetzt hat sich das vierte Verfahren mit einer spezielle 3-Streifen-Kamera mit Strahlenteilung

7 Vgl. Marschall Farbe im Kino, 2005, S. 23.

8 Vgl. Marschall Farbe im Kino, 2005, S. 24.

9 Marschall Farbe im Kino, 2005, S. 25.

10 Vgl. Marschall Farbe im Kino, 2005, S. 26.

11 Dabei wird auf herkömmliches 35mm Filmmaterial mittels bestimmter Optiken das 35mm Bild gequetscht um es in der Projektion wieder in die breite Form zu projizieren.

wegen sinkender Besucherzahlen zu neuen Techniken wie Stereoton, aber auch erstmals zu dem Breitwand-Verfahren wie z.B. das anamorphotische „Cinemascope“-Format. Dies sollte wieder mehr Zuschauer in die Kinos locken, um sich vom normalen damaligen Fernsehformat 4:3 abzuheben¹². Es entstanden im Laufe der Zeit eine Vielzahl ähnlicher Formate wie VistaVision, Techniscope, Super Panavision 70, Ultra Panavision 70 und das Cinerama. Welches Format gewählt wurde hing damals von ästhetischen, ökonomischen aber auch technischen Faktoren ab. Drei Formate setzten sich jedoch durch und werden heute allgemein verwendet. Das 1:1,85-Format, welches hauptsächlich in amerikanischen Produktionen verwendet wird, das 1:1,66-Format, welches mehr in Europa verwendet wird und das 1:1,75-Format. Den erzählerischen Nachteil den das SD-Fernsehen mit seinem Bildseitenverhältnis von 1:1,33 gegenüber den Breitwand-Filmformaten hat, ist optisch deutlich erkennbar. Wie in der Abbildung 1 zu erkennen ist, besteht ein horizontaler Verlust der kompositorischen Elemente, was die Spannung des Bildes oft zerstört¹³.



Abbildung 1: Vergleich der Umwandlung von Film in ein Fernsehbild

2.4 Entwicklung des 3D-Stereofilms

Die Stereoskopie ist eine Entwicklung die über Jahrzehnte bzw. Jahrhunderte ange-dauert hat, bis sie zu dem wurde was wir heute unter 3D-Stereofilm kennen. Da ein umfassender geschichtlicher Abriss bez. der Entstehung des 3D-Films den Rahmen dieses Artikels sprengen würde, wird an dieser Stelle nur speziell auf den letzten Boom in den 1950er Jahre eingegangen. Für weiterführende Informationen bezüglich der historischen Thematik der Stereoskopie sind die Bücher: „The history of stereoscopy“ von ROBERT A. CRONE und „Stereoscopic Cinema and the Orgins 3D-Film“ von RAY ZONE zu empfehlen. In beiden Büchern sind alle notwendigen Informationen zu finden.

12 Vgl. Katz Shot by Shot: Die richtige Einstellung, 2004, S. 443.

13 Vgl. Katz Shot by Shot: Die richtige Einstellung, 2004, S. 444.

Wie bereits erwähnt, sorgte Anfang der 50er Jahre die Verbreitung des Fernsehens zu einem zunehmenden Besucherschwund in den Kinos. Die Filmbranche und die Kinobetreiber machten sich erhebliche Sorgen um ihre Zukunft. Dies führte dazu, dass sich – parallel zum Breitwandfilm- die Aufmerksamkeit der Produktionsstudios wieder der stereoskopischen Kinematographie zuwendete.

MILTON und JULIAN GUNZBURG hatten sich z.B. zusammen mit ARCH OBLER das Ziel gesetzt einen 3D-Film zu produzieren. Daraus entstand im Juni 1952 der Film BWANA DEVILS. Zwei Tage nach deren Veröffentlichung spielte der Film, der nur in zwei Kinos aufgeführt worden war, insgesamt 100 000 Dollar ein - bei Produktionskosten von nur 10000 Dollar. Man kann also behaupten, dass diese Produktion zum damaligen Zeitpunkt ein voller Erfolg war¹⁴. Die Studios und Kinobetreiber zögerten daraufhin nicht lange und erhöhten die Produktion von 3D-Stereofilmen. Seitens der großen US-Studios, die schnell Inhalte produzieren wollten, entstanden die unterschiedlichsten Kamerasysteme, die für unerfahrene Kameramänner eine große Herausforderung darstellten. Diese Unerfahrenheit trug leider dazu bei, dass jede Menge Fehler in der Aufnahme sowie der Projektion entstanden. Weiterhin wurden keine geeigneten stereografischen Filmkonzepte entwickelt, sondern beliebige Inhalte schnell in Massen produziert, um so viel wie möglich von der damaligen Aufmerksamkeit zu profitieren. Kinobetreiber trugen ebenfalls ihren Anteil dazu bei, in dem sie versuchten bei den Projektoren und an den Leinwänden zu sparen¹⁵. All diese Probleme führten zu einer wachsenden Ablehnung der Zuschauer - insbesondere aufgrund der schlechten Qualität der Stereobilder, welche oft Kopfschmerzen und Übelkeit verursachten - und sorgten letztendlich dafür, dass sich der 3D-Stereofilm nicht lange hielt und der Boom im Jahre 1955 schon wieder vorüber war.

Erst ab 2005 führte die Entwicklung des digitalen Kinos dazu, dass sich die Filmindustrie wieder massiv einer neuen visuellen Technik widmete: der stereoskopischen 3D-Produktionstechnik¹⁶ (abgekürzt: S3D). Sowohl alle bekannten Hollywood Studios als auch namhafte Regisseure, wie JAMES CAMERON, der mit AVATAR den erfolgreichsten 3D-Stereofilm in die Kinos gebracht hat, beteiligen sich früh an den neuen S3D-Überlegungen. Traditionelle analoge Produzenten wie GEORGE LUCAS und STEVEN SPIELBERG widmeten sich daraufhin ebenfalls aktiv der neuen Technik. Unter der Leitung von STEVEN SPIELBERG, JEFFREY KATZENBERG und DAVID GEFEN entstanden bei Dreamworks Animation Studios zunächst erfolgreiche stereoskopische Animationsfilme.

KATZENBERG geht aktuell noch einen Schritt weiter und behauptet dass die In-

¹⁴ Vgl. Zone Stereoskopie Cinema and the Orgins of 3-D Film, 2007, S. 184.

¹⁵ Vgl. Lipton Foundations of the stereoscopic cinema, 1982, S. 45.

¹⁶ Anm.: Der Vollständigkeit halber sei darauf hingewiesen, dass der Paradigmenwechsel im digitalen Vertriebssystem zugunsten der Studios sicherlich ebenfalls einen Beschleunigungsfaktor zur Einführung des digitalen Kinos darstellte.

novation des 3D-Stereofilms mit einer Revolution gleichzusetzen ist. Er vergleicht die Tragweite mit der Erfindung des Farbfilms. Für ihn ist der 3D-Stereofilm nicht nur eine Möglichkeit Zuschauer in die Kinos zu locken, sondern er erlaubt es, den Zuschauer zu einem Teil eines immersiven Erlebnisses werden zu lassen. KATZENBERG sieht die Zukunft nicht nur in stereoskopischen 3D-Animationsfilmen sondern schließt den Stereo-Realfilm in diese Entwicklung mit ein mit der Begründung es könnte hierbei ein echter Mehrwert zum bisherigen Filmerlebnis geschaffen werden. Dazu müssten allerdings die Inhalte passen und Filmemacher sich ernsthaft mit den neuen Gestaltungs- und Wirkungsmöglichkeiten befassen¹⁷.

17 Interview: Jeffrey Katzenberg: <http://www.independent.co.uk/arts-entertainment/films/features/the-big-question-are-we-heading-towards-a-3d-screen-future-1868356.html>

3 Grundkonzepte der Filmwahrnehmung und Filmpsychologie

3.1 Filmerlebnis

Ziel aller filmischen Entwicklungen ist es, den Zuschauer mit all seinen Sinnen komplett in das Kinoerlebnis einzugliedern. Psychologen sprechen beim Kinobesuch gerne von einer Traumwelt, ein „emotionales Erlebnis“ in das man gemeinsam eintaucht und auch wieder erwacht¹⁸. FAULSTICH sieht den Kinofilm als „Identifikation und Projektion“ der eigenen Persönlichkeit mit all ihren verdrängten Ängsten. Der Zuschauer wird für eine kurze Zeit in die Geschichte hineingezogen und verlässt sie am Ende der Vorstellung auch schnell wieder. Über den Ablauf hat der Zuschauer dabei wenig kognitive Kontrolle. Rückblickend kann er sich an einzelne Details oft nicht mehr erinnern wenn er den Saal verlässt. Es beginnt und endet wie in einem Traum und das Ergebnis ist mehrheitlich davon abhängig wie stark die emotionalen Verbindung zum Gesehen selbst war¹⁹.

3.2 Dynamik der menschlichen Wahrnehmung

Einen Film mit beiden Augen zu sehen ist eigentlich nichts besonderes, einen Film für beide Augen zu produzieren schon. Die Augen sind jedoch nur ein Teil eines komplexen Systems der Wahrnehmung. Dieses System ist im ständigen Wandel und verändert sich dynamisch²⁰. Mit andern Worten „Es gibt nirgendwo einen wirklichen Stillstand, weder beim Sehen, noch beim Wahrnehmen, noch beim Denken“²¹. Allerdings hat jeder Vorgang seine Wichtigkeit und klare Grenzen. Ist man sich dessen bewusst welche Grenzen es in der menschlichen Wahrnehmung gibt, können weitere Schritte unternommen werden diese Grenzen kennenzulernen und auszuloten.

Die menschliche Wahrnehmung ist ein Prozess, der sich weitestgehend durch das Erlernte, aber auch durch angeborene Faktoren definiert. „Generell herrscht heute die Ansicht vor, dass sich Wahrnehmen und Handeln zusammen ontogenetisch entwickeln“²². SCHNELL umreißt diese Annahme und bindet sie an einige Faktoren wie Wissen, Gedächtnis, Erfahrung, Umweltorientierung und Handlungsdisposition.

Darüber hinaus wird der Zugang zu diesen Faktoren als zyklisch und erweiterbar beschrieben. Die Wahrnehmung unterliegt also einem Prozess, der sich dynamisch erweitert und je nach Motivation verändert werden kann. Diese Tatsache ist entscheidend für die Produktion von Inhalten, denn jeder einzelne Faktor bestimmt den sub-

18 Vgl. W. Faulstich Grundkurs Filmanalyse, 2008, S. 22.

19 Vgl. W. Faulstich Grundkurs Filmanalyse, 2008, S. 23.

20 Vgl. Schnell Medienästetik - Zu Geschichte und Theorie audiovisueller Wahrnehmungsformen, 2000, S. 14.

21 Steinborn Das audiovisuelle Bild und das Wirkliche, 1996, S. 150.

22 Vgl. Guski Wahrnehmen - Ein Lehrbuch, 1996, S. 111.

jektiven Eindruck eines Films. Jede Erfahrung oder gar jedes Wissen, das während oder nach einer Rezeption von Inhalten erworben wird, verändert die Wahrnehmung des Aktuellen und des Zukünftigen²³. Die Wahrnehmung von Film besteht aus Regeln und Erfahrungswerten. Im Laufe einer Rezeption bilden sich neue Erfahrungswerte heraus und verschmelzen mit bereits bestehenden Erfahrungen, die während des gesamten Lebens gesammelt wurden. Diese Fähigkeit des Menschen, Filme zu lesen und Zusammenhänge zu erkennen, bildete sich durch ein bestehendes System von formalen Restriktionen heraus. In diesem filmischen System lebt der Mensch aber auch der Film. Ein Film kann durch sein gesamtes formales filmisches Gebilde ebenfalls eine eigene Welt erschaffen, die sich grundsätzlich von realen Restriktionen unterscheidet. Das Ziel ist es, den Zuschauer von der neu erschaffenen Realität zu überzeugen um diese dann mit seinen Erwartungs- und Erfahrungswerten abzugleichen²⁴.

Regeln des Films können also durch gewisse Filmtechniken gebildet werden, die es dem Zuschauer möglich machen seine Erwartungen zu erfüllen. Sie müssen aber nicht immer zwangsläufig erfüllt werden - auch ein Abgleich von Erwartungen kann eine Zufriedenheit beim Zuschauer erzeugen. Am deutlichsten wird dies in der Montage: Eine Szene beginnt zum Beispiel aus einer „Totalen“ zweier Darsteller. Nun erwartet man im Umschnitt eine nähere Einstellung. Ebenso verhält es sich wenn einer der beiden Darsteller die Szene verlässt, wodurch die Erwartungshaltung entsteht, diesem Darsteller zu folgen. Daher greifen hier wohl die menschlichen Erwartungswerte als Abgleich für die filmische Realität. Hier könnte von einer Art „Grammatik“ gesprochen werden, welche die Wahrnehmung als Formgattung nutzt um einige Elemente des Film zu beschreiben. Im Laufe der Zeit wurden diese Regeln von der Filmtechnik adaptiert und auch tatsächlich im Sinne einer Grammatik bewusst genutzt²⁵.

Dabei steht die menschliche Filmwahrnehmung im Hinblick auf die Stereoskopie vor einer Herausforderung, weil sie noch nicht auf ausprägte Erfahrungswerte aufbauen kann und der wahrnehmungspsychologische Korridor äußerst schmal ist. SCHNELL vergleicht diese Störung tradierter Wahrnehmungsmuster mit der Entstehung neuer „Erfahrungen des Fremdartigen und Befremdeten, der Schnelligkeit und der Beschleunigung, schwindelerregender Sogwirkungen, wie sie von einer Dynamisierung filmischer Bilder ..., in 3D Projektionen beispielsweise ... [sei eine, Anm. d. Verf.] gesteigerte Geistesgegenwart notwendig ... um die Chokwirkung des Films aufzufangen. [Dazu ist das, Anm. d. Verf.] visuelle Erlernen von optischer Beweglichkeit ebenfalls erforderlich.“²⁶ Der Mensch muss sich an diese dynamische Wahrnehmung

23 Vgl. Schnell Medienästetik - Zu Geschichte und Theorie audiovisueller Wahrnehmungsformen, 2000, S. 18.

24 Vgl. David Bordwell Film Art: An Introduction, 2003, S. 53.

25 Vgl. David Bordwell Film Art: An Introduction, 2003, S. 391.

26 Vgl. Schnell Medienästetik - Zu Geschichte und Theorie audiovisueller Wahrnehmungsformen, 2000, S. 13.

von filmischen Formen erst gewöhnen und das benötigt zunächst erst mal Zeit.

Die menschliche Wahrnehmung unterliegt demnach Grenzen bez. der zeitlichen Adaption von filmischen Vorgängen. Wichtig ist hierbei zu verstehen, dass die „Wirklichkeit“ in der sich ein Film bewegt einer Entwicklung unterliegt, die ihre Zeit braucht bis sich Erfahrungswerte bilden und zu Erwartungswerten werden.²⁷

3.3 Wahrnehmungsgrenzen

Wahrnehmungsgrenzen sind in vielen Abschnitten des visuellen Systems zu finden. Das Auge - als Teil dieses Systems - ist alleine schon mehrfach durch seine motorischen und sensorischen Fähigkeiten sowie durch seine Trägheit begrenzt. Es dient zum entdecken und zum aufnehmen von Objekten. Der andere Bereich der Wahrnehmung organisiert sich selbstständig im Gehirn, bestehend auf der Grundlage des „Wir sehen aber wir schauen nicht, wir registrieren, doch wir erkennen nicht, wir nehmen auf - aber nicht wahr“ und besitzt folglich ebenfalls eine gewisse Trägheit. Diese kognitive Trägheit wirkt sich nach SCHNELL insbesondere in zwei Bereichen aus: der „Wahrnehmungsarmut“ und dem „Schock“. Grundsätzlich ist der erste Bereich in den zurückliegenden Lernprozessen verankert und beschränkt die Wahrnehmungsfähigkeit in eine wie SCHNELL es nennt „gedächtnisgestützte Wahrnehmungsarmut“



Abbildung 2: Figuren von Marvin L. Misky und Seymour A. Pappert

Wie sich in einem Testbild von MISKY/PAPPERT zeigt (Abbildung 2) ist für einen Menschen zunächst schwer zu erkennen, ob die Objekte aus einer durchgehenden Linie bestehen oder aus zwei unterschiedlich in sich geschlossenen Objekten. Das Gehirn ist erst nach einer bewussten Überwindung der kognitiven Trägheit den neuen Eindrücken ungesichert zugänglich, es benötigt dann allerdings viel Zeit und Konzentration die Gebilde zu erforschen. Dies bedeutet das der Mensch z.B. in Mediensituationen nicht in der Lage ist, feine Unterschiede ohne „kognitive Konzentration“ zu erkennen.

Der zweite Bereich der Trägheit wirkt sich direkt körperlich auf den Zuschauer aus. Es ist der Effekt des Schocks - häufig ausgelöst von einem unerwarteten Reiz des

²⁷ Vgl. Schnell Medienästetik - Zu Geschichte und Theorie audiovisueller Wahrnehmungsformen, 2000, S. 13.

Wahrnehmungsapparats. Bei einem Schock können Defensivreaktionen oder Abwehrreaktionen die Folge sein, die die Wahrnehmung beeinträchtigen.

Zum Beispiel wenn ein heller Lichtstrahl von der Leinwand auf den Zuschauer trifft und dieser sich dann die Hand vor das Gesicht hält um sich zu schützen oder bei einem sehr lauten Ton bei dem der Zuschauer muskulär zusammen zuckt²⁸.

Der Zuschauer benötigt folglich eine gewisse Zeit um die Dinge wieder bewusst zu erkennen. Diese Wahrnehmungsschwäche wird aktiv im Film genutzt - häufig wird damit sogar ganz bewusst gespielt.

3.4 Filmwahrnehmung und Informationsdichte

Jeder Zuschauer ist bei der Filmbetrachtung mit einem gezielten Maß an Informationsdichte konfrontiert. Diese Informationsdichte erlaubt es dem Zuschauer die Erzählung zu verstehen. Das visuelle System des Menschen ist darauf ausgelegt nur bestimmte Informationen aufzunehmen. Müsste das System alle Informationen aufnehmen würde es schnell überfordert werden²⁹.

Dies geschieht im Film klassisch durch die gezielte Fokussierung der Aufmerksamkeit des Zuschauers auf bestimmte Personen und Objekte durch Form, Farbe, Lichtsetzung oder Bildschärfe. Man spricht hierbei von Bildung einer „selektive Aufmerksamkeit“. Diese beschreibt, welche Informationen konkret aufgenommen und welche ignoriert werden sollen. Wichtig ist dabei zu wissen wie und wann eine Selektion stattfindet. Aufmerksamkeit hängt nicht immer unbedingt von einem visuellen Reiz ab. Aufmerksamkeit hat auch eine mentale Seite. Häufig wird daher das Beispiel „Lesen“ genutzt, da das Lesen eines Buches die Aufmerksamkeit des Lesenden benötigt, dieser aber auch gedanklich das Gelesene verarbeiten muss. Dazu ist Konzentration notwendig, die durch Ablenkungen wie abschweifende Gedanken oder andere äußerliche Reize schnell verloren geht³⁰.

3.5 Informationsverarbeitung im Film

Ein Kinobesucher durchläuft bei der Betrachtung des Films einen Prozess der „kognitiven Psychologie“. Es handelt sich dabei um die „Aufnahme von Informationen, deren Abspeicherung im Gedächtnis und deren Wiederabruf aus dem Gedächtnis“³¹.

Das im Folgenden vorgestellte Modell von PETER OHLER versucht genau diesen

28 Vgl. Schnell Medienästetik - Zu Geschichte und Theorie audiovisueller Wahrnehmungsformen, 2000, S. 14.

29 Vgl. Goldstein Wahrnehmungspsychologie - Der Grundkurs, 2008, S. 132.

30 Vgl. Goldstein Wahrnehmungspsychologie - Der Grundkurs, 2008, S. 133.

31 Vgl. Ohler Filmwahrnehmung, 1990, S. 45.

Prozess der Filmverarbeitung abzugrenzen und zu begründen. Allerdings hängt dieses Modell von mehreren Faktoren ab. Wie im vorangegangenen Abschnitt „Dynamik der Wahrnehmung“ beschrieben, ist vorhandenes Wissen (z.B. das „narrative Wissen“ über Erzählweisen)³² als eigenständiger Faktor dabei zunächst essentiell.

Weiterer Faktor ist das „generelle Weltwissen“, welches der Rezipient benötigt um aus den gezeigten Zeit- und Raumsprüngen logische Schlussfolgerungen zu ziehen³³. Folgendes Beispiel soll dies verdeutlichen: Eine Person läuft auf eine Tür zu, öffnet diese und geht hindurch. Die Tür ist vermutlich der Eingang zu einem Café. In der nächsten Szene befindet sich die Person anschließend im Café. Es ist nicht notwendig dem Zuschauer zu zeigen, dass die Person durch die Eingangstür gegangen ist. Es wird davon ausgegangen, dass man in der nächsten Szene die Person im Café steht und dass dieses Wissen vom Filmschaffenden postuliert wurde.

Ein weiterer und letzter Faktor beschreibt angelernte Wissensbestände des Zuschauers über „filmische Darbietungsformen“ wie Schnitte, Einstellgrößen, Fahrten, Schwenks etc. Der konstruktivistische Ansatz der Informationsverarbeitung von Filmen geht davon aus, dass der Rezipient im Laufe seines Lebens eine Fülle an Informationen über Darbietungsformen erworben hat, was dazu führt, dass bei der nächsten filmischen Informationsaufnahme gewisse Erwartungen entstehen³⁴.

Diese Erwartungen des Zuschauers aus seinem Weltwissen, seinem narrativen Wissen und seinem Wissen über filmische Darbietungsformen sind dynamisch und ändern sich ständig durch den Konsum weiterer Inhalte. OHLER spricht hierbei von einer Art „Hypothesengenerator“³⁵ der erweiterbar ist und den jeder Rezipient während seines filmischen Informationsverarbeitungsprozesses aktiv nutzt.

3.6 Emotional bindende Reizgefühle

CHRISTIAN MIKUNDA schreibt in seinem Beitrag „Kino spüren“ für die GFF Tagung 1989 über den „Orientierungsreflex“ und „Orientierungsreaktion“ des Zuschauers bei der Betrachtung von Filmen, welche beide zu einer starken emotionalen Bindung an den Film führen. Sein Ausgangsgedanke ist Folgender: Jeder Reizwechsel im audiovisuellen Medium sorgt generell für ein bestimmtes Aktivierungsgefühl beim Rezipienten. „Die so genannten afferenten Neuronen in der Hirnrinde reagie-

32 Vgl. Ohler Filmwahrnehmung, 1990, S. 47.

33 Vgl. Knut Hickethier Filmwahrnehmung, 1990, S. 48.

34 Vgl. Ohler Filmwahrnehmung, 1990, S. 48.

35 Der Hypothesengenerator, enthält die aufbereiteten und vorverarbeiteten Wahrnehmungselementen. Er verarbeitet und verändert diese unter Verwendung von Regeln wie Gestaltungsgesetze sowie der Nutzung von bestehendem Wissen und gibt ein Percept aus, Perzeption und kognitive Verarbeitung der Sprechmelodie - Oliver Niebuhr

ren auf Sinnesreize mit neuroelektronischen Impulsen”³⁶. Daraus ergibt sich für ihn, dass ein rasch auftretender Sinnesreiz, der besonders schnell und unerwartet im Film auftritt, zu einem körperaktivierenden Impuls führt, der die „physische Erregung“ des Zuschauers steigert. Allerdings ist dieser körperliche Reflex nicht in unendlicher Frequenz wiederholbar, da sich der Rezipient irgendwann auf den Reiz einstellt und einen Lernprozess durchlebt. Dies führt dazu, dass der Effekt irgendwann nur noch eine schwache bis gar keine Reaktion hervorruft, die emotionale Bindung an den Film nimmt deutlich ab. Deswegen unterscheidet MIKUNDA auch zwischen „Orientierungsreflex“ und „Orientierungsreaktion“. Die Orientierungsreaktion geht als Gegenkonzept aus einem „bedeutungsvollen Reiz“ für den Zuschauer hervor und klingt in ihrer Intensität eher langsam ab³⁷.

Diese beiden Konzepte können wichtige Hinweise für die Planung und Durchführung von Szenen und Folgeszenen beim 3D-Stereofilm geben. Kurz gesagt bedeutet dies zunächst, dass ein 3D-Effekt irgendwann seine Wirkung verliert, wenn er zu häufig auf die gleiche Weise präsentiert wird.

Desweiteren kann daraus gefolgert werden, dass neue Orientierungsreflexe im 3D-Stereofilm in ihrer emotionalen Wirkung und Bindung beachtet werden müssen beispielsweise Defensivreaktionen wie Zusammenzucken, Ducken oder Ausweichen³⁸. Diese entstehen, wenn „Wahrnehmungsobjekte“ mit einer hohen Geschwindigkeit oder schnell ansteigenden Intensität auf den Rezipienten treffen.

In letzter Konsequenz bedeutet dieser Denkansatz auch, dass eine langanhaltende emotionale Bindung des Zuschauers durch Stereo-3D nur erzielt werden kann, wenn die Stereografie zu einem narrativ-bedeutungsvollen Tiefenreiz verwendet wird.

3.7 Stereofilm-Benefits: Immersion oder Theatralik?

Es stellt sich abschließend die Frage, welcher filmpsychologische Nutzen von einer gesteigerten Tiefenwahrnehmung des Zuschauers im Film zu erwarten ist.

ROBIN CURTIS beschreibt die Tiefenwahrnehmung des Zuschauers als „synästhetischen Erlebnis“- als „Einfühlung in den Raum“. Es entsteht letztendlich eine „virtuelle Tiefe“, basierend auf den eigenen Erfahrungen, dem angesammelten Wissen und der eigenen Vorstellungskraft des Zuschauers von Tiefe³⁹.

Zusätzlich scheint es wichtig sein zu klären, ob in der Summe aller Gestaltungs- und Inszenierungsmöglichkeiten eine Tiefenwahrnehmung zum Zuschauer hin oder vom

36 Mikunda Filmwahrnehmung, 1990, S. 90

37 Guski Wahrnehmen - Ein Lehrbuch, 1996, S. 115.

38 Vgl. Guski Wahrnehmen - Ein Lehrbuch, 1996, S. 115.

39 Vgl. Curtis Montage-AV 2009, 2009.

Zuschauer weg entsteht. Grund: Der Kunsthistoriker MICHAEL FRIED definiert den Begriff Immersion als „imaginäres leibliches Eintauchen in die Atmosphäre... [wenn, Anm. d. Verf.] sie von den dargestellten Objekten und Figurationen“⁴⁰ entsteht. Der Zuschauer lässt sich demnach besonders gut in eine Szene hineinziehen, wenn die Darsteller von ihrer Seite aus etwas absorbieren. Die Theatralik hingegen ist eine eher minimalistische Ausdrucksform - allein auf die Gewinnung von Aufmerksamkeit begrenzt. Werkzeuge sind hier meist starke Reize. Für den Zuschauer hat dies Folgen: er fühlt sich vom betrachteten Geschehen nicht hineingezogen sondern die Handlung springt ihn sozusagen an. Folglich unterscheiden sich Immersion und Theatralik grundsätzlich in der psychologischen Wirkung auf den Zuschauer⁴¹.

Aus der Sicht der Zuschauer liegt also die wesentliche Berechtigung des Stereofilms in seiner Immersionswirkung. Die Zuschauer werden besser in das dramaturgische Geschehen eingebettet und bauen im besten Fall ein verstärktes (emotionales) Verständnis zum Film und seiner Botschaft auf.

40 Vgl. Curtis Montage-AV 2009, 2009.

41 Vgl. Voss Montage-AV 2 [2008], 2008.

4. Bildsprache - Definition und Einordnung

4.1 Bildsprache - Definition

Der Begriff der Bildsprache leitet sich aus Elementen des Bildes ab, um sie mit einem System beschreibbar zu machen, welches an der menschlichen Sprache angelegt ist. Dabei wird sich oft des Ansatzes von CHRISTIAN METZ der sog. Filmsemiotik bedient. Gemeint ist ein Kommunikationsprozess zwischen Film und Zuschauer mittels Zeichen. Der Signifikant beschreibt dabei das eigentliche „Wesen“ des Zeichens und der Signifikat dessen „Bedeutung“⁴². In der sprachlichen Kommunikation zwischen Menschen stehen Signifikant und Signifikat für „Klang und Bedeutung“ und für einen Kanal, der bidirektional aufgebaut ist und einer großen Anzahl an Faktoren, die diesen Kanal beeinflussen können. Beim Film hingegen ist die Kommunikation klar geregelt und zunächst unidirektional. Signifikant und Signifikat sind gleich, das Gesehene wird unmittelbar verstanden⁴³.

Man spricht folglich von der Semiotik des Films, die aufgrund der Leichtigkeit des Verständnisses von Film trotzdem schwer zu erklären ist⁴⁴. Daraus ergibt sich eine kuriose Asymmetrie: Aus Sicht der Produktionsseite gibt es nämlich eine Fülle an Möglichkeiten ein Bild darzustellen – aus Sicht der Rezipientenseite besteht jedoch nur eine Möglichkeit, dieses ausgegebene Bild zu verstehen. Es ist folglich für den Rezipient essentiell, das Medium richtig lesen zu lernen.

„Je besser man Bilder liest, desto besser versteht man sie und umso mehr Macht hat man darüber. Der Leser einer Buchseite erfindet das Bild, der Leser eines Films tut das nicht“⁴⁵.

METZ spricht in seiner Theorie von einer „kinematographischen Grammatik“, anhand derer er versucht die komplexe Struktur einer Filmhandlung zu erklären. Allerdings beschreibt er dieses Geflecht als „offenes“ und „organisches System“ welches sich permanent weiterentwickelt.

Der Sprachwissenschaftler JAN-MARIE PETERS nimmt an, dass die Sprache des Films einem linguistischen System gleichzustellen ist. Er geht von der Annahme aus, dass mit der Sprache bestimmte Gefühle, Gedanken oder auch Wünsche mit Hilfe von „Zeichen“ ausdrücken können und überträgt diese Annahme auch auf den Film. Er sieht das Zeichen als ein Mittel an, das Verstanden werden kann und geht davon aus

42 Vgl. Schnell Medienästhetik - Zu Geschichte und Theorie audiovisueller Wahrnehmungsformen, 2000, S. 171.

43 Vgl. Schnell Medienästhetik - Zu Geschichte und Theorie audiovisueller Wahrnehmungsformen, 2000, S. 173.

44 Vgl. Schnell Medienästhetik - Zu Geschichte und Theorie audiovisueller Wahrnehmungsformen, 2000, S. 172.

45 Monaco Film Verstehen, 2007, S. 173.

dass der Träger hierbei nicht so wichtig ist.

„Da die Zeichenelemente Wort und Bild gleichermaßen einen Abstraktionsgrad im Verhältnis zu der von Ihnen jeweils „bezeichneten“ Sache aufweisen, lassen sie sich übereinstimmend begreifen als Zeichen, in dem das Abgebildete verstanden wird.“⁴⁶

Eine Gruppe von Bildern ist somit eine Ansammlung von sinnhaltigen Informationen, die eine Mitteilung erzeugen, die dem Publikum dazu dient den Film zu verstehen. SCHNELL spricht hierbei von einem „konstruierten Bild“ das sich als „syntaktisches, respektive morphologische System“ versteht und das eine eigene „Sprachsystematik“ besitzt. Die notwendigen Strukturelemente bilden Montage und Inszenierung, sie dienen als „Filmischer Satzbau“⁴⁷.

Der von SCHNELL dargelegte Vergleich mit der Sprache ist schwierig und wird oft kritisiert, da das Wort nicht mit einem Bild verglichen werden kann. Im Film hingegen ist das Bild durch das Zusammenstellen in Szenen als solches unendlich erweiterbar. Ein Wort besitzt dagegen durchaus seine Eigenständigkeit und hat eine klare Bedeutung und ist somit in seiner Nutzung stark eingeschränkt. „Nicht weil das Kino eine Sprache ist, kann es uns so schöne Geschichten erzählen, sondern weil sie es uns erzählt hat, ist es zu einer Sprache geworden.“⁴⁸

Abschließend sei aber festgehalten, dass der Begriff Bildsprache des Films mehr umfasst als die visuellen Gestaltungsweisen (visual style) im Sinne von cinematografischen Codes (auf die wir später eingehen) sondern den gesamten Kommunikationsprozess zwischen Filmschaffenden, dem Werk und seinen Zuschauern beschreibt. In diesem Sinne soll der Begriff „Bildsprache von 3D-Produktionen“ auch im Weiteren verwendet und näher betrachtet werden.

4.2 Denotation und Konnotation

Es ist also möglich, dass ein Film mit seinen sofort verstehbaren Bildern eine Bedeutung transportiert. Daher wird in der Filmsemiotik zwischen „Denotation“ und „Konnotation“ unterschieden. Denotation meint den tatsächlichen unumstößlichen Inhalt oder die Form des Zeichens. Jedes Zeichen transportiert aber auch eine sinnhaltige Bedeutung, die für den Empfänger zu einer subjektiven, konnotativen Botschaft wird. Zur Verdeutlichung wird immer gerne das Bild einer „weißen Taube“ zitiert. Die Denotation ist das Abbild eines weißen Vogels aus der Familie der Columbidae.

46 Schnell Medienästetik - Zu Geschichte und Theorie audiovisueller Wahrnehmungsformen, 2000, S. 172.

47 Vgl. Schnell Medienästetik - Zu Geschichte und Theorie audiovisueller Wahrnehmungsformen, 2000, S. 173.

48 Vgl. Metz Semilogie des Films, 1972, S. 73.

Die Konnotation dieses Bildes ist (in der westlich geprägten Welt) aber „Frieden“. In anderen Kulturen könnte die konnotative Bedeutung eine deutlich andere sein.

Jenseits dieses einfachen Kulturwissens (z.B. als Drehbuch-Vorgabe) entstehen im Film zusätzlich konnotativen Faktoren bei der Erzeugung des Bildes, da die Ersteller mit der Wahl der Location, der Lichtstimmung, der Schauspieler, den Kameraeinstellungen und mit Farbe oder Kontrast das Bild ebenfalls konnotativ beeinflussen⁴⁹.

In letzter Konsequenz muss das Erfassen von filmischen Konnotationen auch davon abhängig gemacht werden, welche Aufnahmen aus einer Vielzahl von Einstellungen für die Montage ausgewählt und zusammengesetzt wurden. Allein aus dieser Auswahl ergeben unterschiedliche Bedeutungen. Wird etwas bewusst weggelassen, was von Bedeutung ist, wird die gesamte Wahrnehmung der Szene verändert. Dies wird auch als „paradigmatische Konnotation“ bezeichnet. Das Gegenstück dazu die „syntagmatische Konnotation“ beschreibt die Anordnung der Aufnahmen im gesamten Film und wie sie in Beziehung zueinander stehen⁵⁰.

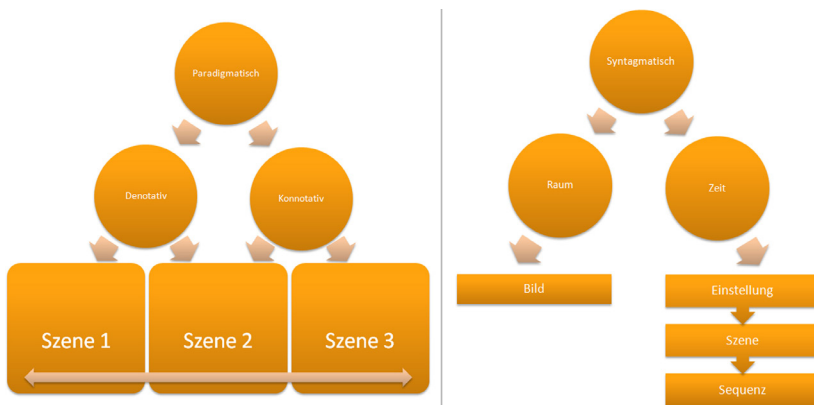


Abbildung 3: Paradigmatische und syntagmatische Konnotationsprozesse der Filmerstellung

Die paradigmatische und die syntagmatische Konnotation stellen folglich die konzeptionellen und gestalterischen Eckpfeiler des Filmhandwerks dar. Es wird an späterer Stelle zu untersuchen sein, ob und wenn ja, welche konnotativen Einflüsse die Stereografie auf einen Film haben kann.

49 Vgl. Monaco Film Verstehen, 2007, S. 163.

50 Vgl. Monaco Film Verstehen, 2007, S. 164.

4.3 Kinematografische Grammatik und Syntax

Film besteht letztendlich aus der Summe vieler einzelner Teile und nicht nur aus einzelnen Einstellungen oder Szenen. Der Zuschauer begreift und bewertet letztendlich immer den Film als Ganzes. Filmgrammatik meint daher die umfassende, systematische Beschreibung und Abbildung des gesamten filmischen Kommunikationsprozesses von der Erstellung bis zur Rezeption. Ein wesentlicher Punkt in dieser systematischen Beschreibung sind die Regeln zur Verwendung der einzelnen Elemente - gemeinhin als Syntax oder kinematografische Codes bezeichnet. Ähnlich der gesprochenen oder geschriebenen Sprache ist es auch in Filmen wichtig, dass diese Regeln und ihre Abhängigkeiten untereinander bekannt sind und eingehalten werden.

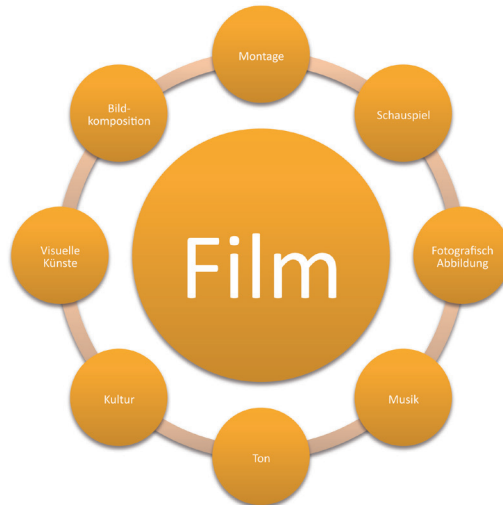


Abbildung 4: Schaubild kinematografischer Codes

Wesentliche kinematografische Codes des klassischen Films sind der fotografische und bildkompositorische Code⁵¹, der musikalische Code, der Schauspiel-Code sowie die Codes der Montage. All diese Regeln haben eines gemeinsam, sie vermitteln in ihren internen Ausprägungen und Beziehungsgeflechten untereinander Konnotationen und geben in ihrer gezielten Komposition dem Film etwas Eigenständiges⁵².

51 Vgl. Kritzberger Multimediale und interaktive Lernräume, 2004, S. 93

52 Vgl. Monaco Film Verstehen, 2007, S. 164.

4.4 Stereo als kinematografischer Code?

Stereografie wird oft spontan als Subcode zum photographischen Code subsummiert. Allerdings wäre auch eine Einordnung zum bildkompositorischen Code möglich (s. Inszenierung über Raum und Zeit) oder in den Code der visuellen Künste.

Die historischen Betrachtungen aus Kapitel 2 legen eher eine Betrachtung der Stereografie als eigenständigen kinematografischen Code nahe. Sie zeigen uns beispielsweise, dass sich tradierte kinematografische Codes oder Subcodes im Zuge von technischen Entwicklungen verändern mussten oder sogar verschwanden. Die Stereografie von vorneherein als Subcode eines bestehenden Codes zu verorten erscheint daher kritisch.

Die hohen systemischen Wechselwirkungen der Stereografie mit den anderen kinematographischen Codes legen ebenfalls nahe, die Stereografie zunächst als eigenständigen, zusätzlichen kinematografischen Code zu betrachten.

Dazu ein typisches Beispiel: Ein häufig verwendeter, tradierter photographischer Subcode ist das „Vorbeiwischen“ von Gegenständen die sich sehr nahe an der Kamera befinden. Dies erzeugt beim Zuschauer einen Reizwechsel⁵³. Genutzt wird dieses Mittel auch als Schnitt um den Ort oder die Zeit zu wechseln, somit hat dieser Subcode eine typische Wechselwirkung mit der Subcodes der Montage. Dieses „Vorbeiwischen“ ist jedoch problematisch im 3D-Stereofilm, denn die Tiefenwirkung bricht aus technischen und wahrnehmungspsychologischen Gründen an diesen Stellen meist unmittelbar zusammen.

53 Knut Hickethier Filmwahrnehmung, 1990, S.89.

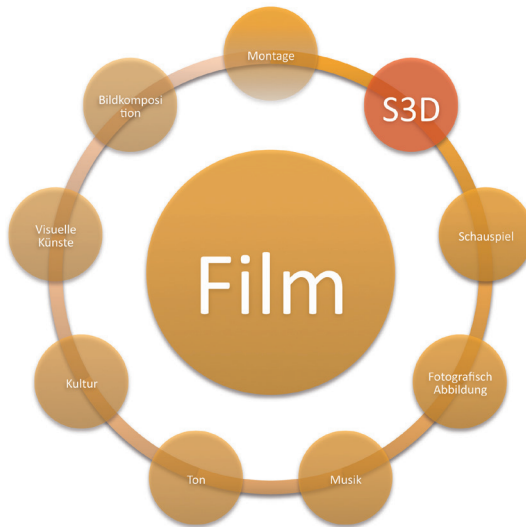


Abbildung 5: Einordnung der Stereografie als neuer kinematografischer Code

Bleibt zum Schluss die Frage, welchen konnotativen Benefit ein eigenständiger kinematografischer Code mit Namen „S3D“ haben könnte?

Die Antwort darauf liegt ist vorangegangen Kapitel „Filmwahrnehmung“ verankert. Dieser neue Code lässt neue Stimmungen beim Betrachten zu. Er kann eine neue, eigenständige Form vom emotionaler Verbundenheit mit dem Film erzeugen im Sinne der verstärkten filmischen Immersion.

5 Bildsprache des 3D-Stereofilms

Zur Definition einer Bildsprache des 3D-Stereofilms gehört folglich die grammatische insbesondere aber die syntaktische Beschreibung aller in ihm angewendeten kinematografischen Codes bzw. Subcodes sowie ihre Beziehungen und Veränderungen untereinander. Im Folgenden werden zunächst nur die Wichtigsten betrachtet.

5.1 Inszenierung

Die Inszenierung ist eine Gestaltungsmöglichkeit einer Szene. Die handelnden Darsteller füllen hierbei einen Raum durch Formen ihrer Inszenierung. Die Darsteller nutzen hierbei Körpersprache, Timing und den zur Verfügung stehenden Raum als Interaktionsbereich der Inszenierung.

Betrachtet man die Geschichte des Kinos in Hinblick auf seine visuellen Gestaltungsweisen, so kann man sagen, dass es stets darauf ankommt, auf sehr unterschiedliche Weisen die Aufmerksamkeit des Zuschauers zu wecken. Diesen somit durchgängig an den Film zu binden und ihm relevante Informationen über die Realität zu übermitteln, die auf der Leinwand gezeigt wird⁵⁴.

Der Film bildet sich aus unterschiedlichsten Räumen heraus, wie Tiefenräume oder die Räume des kompositorischen Bildes. Diese Räume gilt es zu füllen, hierbei werden anhand unterschiedlichster Methoden, Informationen aktiv an den Zuschauer vermittelt.

5.2 Spielraum oder Raumspiel

Filme bestehen aus unterschiedlichsten Filmstilen. Es gibt zum Beispiel Filmstile, die sich eine bestimmte Anordnung der handelnden Personen erlauben. Dies wird häufig durch das Stilmittel im Film gefestigt. Diese Filmstile sind sehr darstellerfixiert und mit einer im Hintergrund laufende Rahmenhandlung besetzt. Sie besteht aus viele Nahaufnahmen der Darsteller das den Film dadurch prägt. Durch den Einsatz dieses Stils werden allerdings andere kinematografische Mittel (Codes) ausgeblendet. Die Räumlichkeit verfällt und dem Zuschauer wird die Möglichkeit verweigert abseits der Darstellungsebene Informationen aufzufassen. Weiterhin kann der Film den Raum auch als Ganzes nutzbar gestalten. Dabei stehen die auftretenden Spannungen, wie der Wechsel von Vordergrund zu Hintergrund und die folglich gebildeten Nebeninformatio- nomen im Fokus. Zusammenfassend kann der Film einen Spielraum erzeugen und dieser ist im Folge dessen nur begrenzt nutzbar oder der Film nutzt den Raum als Mittel um diesen voll auszufüllen. Jede Variante hat ihre Vor- und Nachteile⁵⁵.

⁵⁴ Vgl. D. Bordwell Visual Style in Cinema - Vier Kapitel Filmgeschichte, 2006, S. 13

⁵⁵ Vgl. D. Bordwell Visual Style in Cinema - Vier Kapitel Filmgeschichte, 2006, S. 13.

Was macht den Raum des Schauspiels zum Raum der Handlung. Die Wichtigkeit der Objekte und das Geschehen der handelnden Darsteller, sind das Leitbild einer Inszenierung im Raum. Der Zuschauer erhält eine entscheidende Rolle. Er wird nun vor die Aufgabe gestellt die Handlung zu lesen. Häufig geschieht dies ganz von allein, da sich der Film ganz bewusst kultureller Werte wie Sympathien, Leidenschaft oder Liebe bedient.

Es gibt nun unterschiedlichste Möglichkeiten einen Raum zu füllen. Dieser kann durch seine Tiefe eine Wichtigkeit besitzen. Je nach dem wie die Objekte platziert werden und wie sich im Laufe einer Inszenierung verhalten, verändert sich deren Stellenwert. Mit dem Ziel, entweder dramaturgische Elemente eine Gewichtung zu geben oder ganz pragmatisch einen Schnitt zu vermeiden. Wichtig ist nur, wie viele Information zum Zuschauer gelangen sollen und was der Zuschauer bemerken darf. Die Interpretation der Handlung durch Zuschauers anhand der erhaltenen Informationen spielt ebenfalls eine Rolle⁵⁶. Zum Beispiel kann sich ein Objekt in Form eines Darstellers durch die einzelnen Staffelebenen bewegen, von Vorder-, zu Hintergrund. Dies Bewegung lenkt die Aufmerksamkeit des Zuschauers. Was bedeutet dies jedoch für den Raum? Dieser erhält dadurch ein Gewicht, das je nach Staffellung, Bereiche der Aufmerksamkeit steuert und die Dramaturgie einer Szene unterstützt⁵⁷. Im 3D-Stereofilm bestehen die selben Möglichkeiten, nur das der Raum für die Tiefe besser zur Geltung gebracht werden kann. Ebenfalls unterstützt 3D-Stereo die Gewichtung der Aufmerksamkeit durch seine Bindung an aufgezwungene Tiefenebenen.

5.3 Klassische Organisationsprinzipien für Raum und Zeit

Für jeden Filmemacher stellt sich die Frage wie er was und in welcher Art filmen soll. Aus diesem Grund wurden in der Filmanalyse die Begriffe, "Mise en Scène" und "Montage" gebildet. Beide bilden getrennt von einander die Organisationsprinzipien für Raum und Zeit. Die Mise en Scène beschreibt eine Veränderung im Raum und die Montage eine Veränderung der Zeit. Beide unterstützen den Film um die "psychologische Realität" zu bilden und damit die "physische Realität" zu überschreiten⁵⁸.

5.3.1 Mise en Scène

Der Begriff "Mise en Scène" - "in Szene setzten" beschreibt die Veränderung des Raumes durch die Kamera, die eine besondere Wichtigkeit in Verbindung zur Darstellung eines Schauspiels herstellt. Die Mise en Scène wird oft als statisch angesehen, welches sie nicht unbedingt ist. Sie ist für die Darstellung einer Einstellung mittels Bildkomposition und Bildkardrierung wichtig für jeden Filmemacher⁵⁹.

56 D. Bordwell Visual Style in Cinema - Vier Kapitel Filmgeschichte, 2006, S. 17

57 Vgl. Katz Shot by Shot: Die richtige Einstellung, 2004, S. 316.

58 Vgl. Monaco Film Verstehen, 2007, S. 177.

59 Vgl. Monaco Film Verstehen, 2007, S. 177.

Die Bildkardrierung bestimmt dabei zunächst die Bildkomposition durch der Festlegung des Bildformates. Dadurch werden die ersten Kompositionsmöglichkeiten festgelegt. Bestimmte Filmformate wie "CinemaScope" legen den Fokus eines Bildes auf Hintergrund, da das Bild bei diesem Format stark in die Breite gezogen wird. Damit wird der Hintergrund stärker Teil der Szene. Andere Formate hingegen wie das, 4:3-Format, eignen sich hervorragend für Interview oder Gesprächssituationen. Bei der Bildkardrierung gibt es zwei Varianten die "offene" und "geschlossene" Form. Die geschlossene Form beschreibt Einstellungen die in sich visuell geschlossen sind. Dabei sind die Handlungen im Raum für den Zuschauer immer wahrzunehmen. Die offene Form hingegen beschreibt einen erweiterten Spielraum der Handlungen, die für Zuschauer nicht mehr sichtbar sind. Zum Beispiel wenn eine Handlung den visuellen Raum verlässt und sich außerhalb des sichtbaren Bildes fortsetzt⁶⁰. Die Bildkomposition stellt den Filmemacher vor die Aufgabe den Spielraum zu füllen, indem er ihn in drei Bereiche aufteilt. Wie in der Abbildung 6 zu erkennen ist bestehen die drei Bereiche aus:

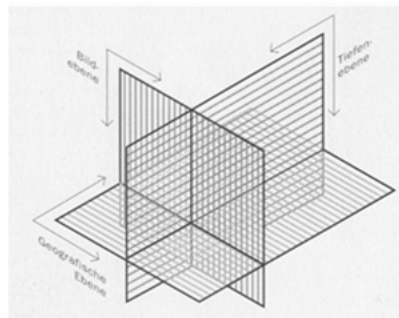


Abbildung 6: Die drei Kompositionsebenen nach Monaco

- Der Bildebene, welche das Bild und deren Handlung bestimmt.
- Der geografische Bereich stellt diesen Raum zur Verfügung.
- Dem Bereich der Tiefen-Wahrnehmung einer Szene.

Wie in der Abbildung 6 zu erkennen ist sind alle drei Bereiche voneinander abhängig und bestimmen die Bildkomposition.

Was genau bedeutet es aber für den Film? Zunächst wird dieser dadurch ein wenig unberechenbarer für den Filmemacher. Das Schauspiel einer Szene wird verlängert, ohne dass ein Schnitt entsteht. Die Szene besteht dadurch aus viel Spontaneität und Improvisationskunst des Darstellers und Interaktionen zwischen den Darstellern. Die

⁶⁰ Vgl. Monaco Film Verstehen, 2007, S. 189.

Mise en Scène wird häufig als zu langweilig empfunden, dabei wird sie wohl häufig missverstanden. Die Methode beschreibt Gleichgewicht und Spannung, Dynamik zwischen den Darstellern und eine realistische Ausdrucksform des Schauspiels⁶¹.

5.3.2 Montage

Der moderne Film besteht aus Bildmontagen, dies macht ihn für den Zuschauer zunächst einfacher zu verstehen. Der Zuschauer soll sich als Beobachter im Raum fühlen und sollte sich die Szene aus so vielen Einstellungen wie Möglich zusammensetzen. Montage gilt im Film als Organisationseinheit, technisch ist es nur ein Schnitt, aber gestalterisch ergibt sich dadurch ein Mittel, das die Qualität der Bildfolge bestimmt. SCHNELL benennt dies als strukturelles Verhältnis zu einander, das vor allem bestimmt ist durch das wie der Film zusammengefügt ist und nicht welche filmischen Elemente verwendet wurden, oder gar wie Schauspiel oder Handlung damit verbunden sind. Dabei kann die Montage hitzige und spannende, aber auch genauso gut ruhige Momente erzeugen. Sie bindet sich an einen Rhythmus, der wiederum verbunden ist mit der Dramaturgie einer Szene. Besonders gut umgesetzt hat dies ein bestimmtes Genre. Der "Action-Film" reizt die Grenzen der Montage nahezu aus und übersteigt fast das Wahrnehmbare⁶². Die Schnittfolge wird dadurch fast das Hauptstilmittel. Die durchschnittliche Einstellungsfrequenz von unter 2,7 Sekunden bei einigen Filmen ist dafür bezeichnend⁶³. Diese hängt natürlich ganz stark mit dem Filmgenre zusammen und der Produktion für den Massenmarkt. 3D-Stereo wird aber nur eins dieser Stilmittel werden, wenn es von dem Massenmarkt akzeptiert und von der Produktionsseite angewandt wird. Bei der Stereoskopie kommt ein weiteres Momentum hinzu, das Gehirn benötigt wesentlich mehr Zeit um das Gesehene zu verarbeiten.

Voraussetzung für ein angenehmes Sehen ist, dass die beiden Bereiche der Konvergenz und Akkomodation nicht sehr weit auseinander liegen⁶⁴.

Schnittfrequenz und Schnittrhythmus

Die Länge einer Einstellung bestimmt die Bedeutung des Bildes, wobei eine Szene nicht unbedingt einem bestimmten Rhythmus folgen muss. Ein Rhythmus und die Einstellungslänge hängt zum einen davon ab, wie man eine Geschichte erzählen möchte, zum anderen welches Stilmittel verwendet wird. Heutzutage wird der Rhythmus als Stilmittel angesehen und sorgt bei Filmen die in ihre dramaturgische Erzählweise viel Hektik einbauen, als das Mittel diese Hektik auch auf den Zuschauer zu übertragen. Dabei ist es wichtig, dass die Wahrnehmung des Zuschauer gelenkt wird, da dieser meist nicht viel Zeit hat, sich auf die Informationen einzustellen.

61 Vgl. D. Bordwell Visual Style in Cinema - Vier Kapitel Filmgeschichte, 2006, S. 113.

62 Vgl. Schnell Medienästetik - Zu Geschichte und Theorie audiovisueller Wahrnehmungsformen, 2000, S. 99.

63 Vgl. D. Bordwell Visual Style in Cinema - Vier Kapitel Filmgeschichte, 2006, S. 171.

64 Vgl. Kuhn Stereo-Fotografie und Raumbild-Projektion, 1999, S. 22.

Essentiell ist dabei, dass die gesamten Szenen perfekt aufeinander abgestimmt ist um die Wahrnehmung in der geringen Zeit zu gewährleisten⁶⁵. Ob das der Stereoskopie zu Gute kommt, ist wegen der Eingewöhnungszeit für jede Szene fraglich? Allerdings wird sich die Wahrnehmung des Zuschauers mit der Zeit darauf einstellen, was dazu führt dass häufiger Filme geben wird die versuchen diese Grenzen auszureizen.

Tiefensprünge

Sprünge in der Tiefe lassen sich kaum vermeiden. Je nach Einstellung und stereoskopischer Ausrichtung kann es hierbei sehr unangenehm für den Zuschauer werden, wenn er sich beim Wechsel der Einstellung an die neue Tiefenausrichtung gewöhnen muss. Es ist notwendig zu wissen wann ein Tiefensprung passiert und welche Intensität der Sprung entwickelt. Es wird hierbei von einer Tiefenfortführung bzw. Continuity gesprochen. Dabei ordnet sich die Tiefe einer Szene, der Restriktion von von starken Tiefensprüngen unter. Hier sollte ein gewisses Maß angesetzt werden, das nicht ausgereizt werden darf. Um diesen Sprüngen entgegen zu wirken, kann man ein sogenanntes "Tiefenscript" erstellen. Auf diese Weise ist es möglich genau kontrollieren wie die einzelnen Tiefenebenen beschrieben sind. Im Abschnitt "Gestaltungsmöglichkeiten beim 3D-Stereofilm" wird beispielhaft ein Tiefenscript genauer beschrieben.

Einstellungsdauer

Die Einstellungsdauer steht unmittelbar in Verbindung mit der Dramaturgie. Ein Schnitt kann als sehr angenehm empfunden werden, wenn dafür Voraussetzungen erfüllt werden. Der Zusammenhang zwischen der Wahrnehmung einer Szenen und die gezeigten Handlungen sind dabei sehr wichtig. Erlangt der Zuschauer ein Verständnis zu dem Gezeigten, wird der Schnitt als sehr angenehm empfunden. Die Einstellungsdauer beschreibt somit die Dauer, einer Information die zum Zuschauer gelangt. Reicht diese nicht aus wird der Schnitt bemerkt und sorgt für einen Bruch in der Wahrnehmungs- und Verarbeitungskette. Die Produktion bevorzugt häufig eher kurze Einstellungen als verstärkte dramaturgische Elemente, dabei wird die notwendige Zeit zur Wahrnehmung eher zweitrangig betrachtet. Eine Alternative für kurze Einstellungsdauer wäre die Inszenierung in der Tiefe das einer der Form der Mise en Scène gleich kommt⁶⁶. Dies würde der Stereoskopie dies entgegen kommen und sollte bei der Umsetzung in Betracht gezogen werden. Es würden den Fokus weg von möglichst starker Montage, hin zu dramaturgischen Bildkompositionen und einem inhaltlich passenden dramaturgischen Ablauf legen. Dieses Elemente der Inszenierung könnten sich wieder vermehrt durchsetzen und die Einstellungsdauer verlängern. Als gutes aktuelles Beispiel ist hier *Inglourious Basterds* von Quentin Tarantino aufzuführen. Die Abbildung 7 beschreibt hierbei das gestalterische Vorhaben. Hier besteht die Anfangsszene aus einer Anzahl von wenigen Sequenzen. Es finden sehr wenige Schnitte statt und selten wird ein Perspektivenwechsel durchgeführt. Das wirkt zwar auf dem Papier sehr statisch vermittelt jedoch den gesamten Szenen viel Dynamik.

65 Vgl. D. Bordwell *Visual Style in Cinema - Vier Kapitel Filmgeschichte*, 2006, S. 22.

66 Vgl. D. Bordwell *Visual Style in Cinema - Vier Kapitel Filmgeschichte*, 2006, S. 174.

Dieser ruhige Aufbau wird erst gegen Ende der Szene durch eine leichte Bewegung der Kamera gestört, dadurch wird eine starke dramaturgische Wirkung erzielt. Die Nahaufnahmen am Ende beschreiben den emotionalen Höhepunkt des Filmabschnitts. Mit relativ einfachen visuellen Mitteln wurde ein spannendes Gebilde geschaffen, bei dem sich der Einsatz von visuellen Mitteln, Schauspiel und Bewegung der Dramaturgie gegenseitig bestimmen und in sich geschlossen bildete.

Die Frage die sich im Hinblick auf die Stereoskopie stellt, ob zum einen diese Szenen auch stereoskopisch funktioniert hätten und zum anderen wie sich die Wirkung mittels der Stereoskopie verändert hätte. Die stilistische Grundlagen wie, Einstellungsdauer, Schnittrhythmus, Schnittfrequenz aber auch kompositorische Bildgestaltung wären für eine stereoskopische Abbildung wahrscheinlich sehr geeignet gewesen.



Abbildung 7: Szenen aus Inglourious Basterds

5.4 Stereo und Filmdauer: Effekt oder Stilmittel

Zunächst gilt es grundsätzlich zu klären was genau ein Effekt ist und welche Definition für einen 3D-Stereofilm gefunden werden kann. Ein Effekt beschreibt sich grundsätzlich in seiner zeitlichen Begrenzung bzw. in der Dauer seines Einsatzes. Ein Effekt kann von kurzer Dauer, also vielleicht wenige Sekunden genutzt werden und dient als unterstützendes Element im Film, genauso wurde bzw. wird die Stereoskopie im Film gesehen, mit seinen Fähigkeiten Dinge räumliche Darzustellen. Dazu gibt es die Definition eines Stilmittels. Ein Stilmittel ist nicht zeitlich begrenzt und verfügt über die Eigenschaft einen Film als Ganzes zu gestalten über eine ganze Episode hinweg, einen Zeitraum oder gar über den gesamten Film nutzt. Außerdem beschreibt es den Film und macht diesen für den Zuschauer lesbar. Ein Stilmittel gibt einem Film eine Bedeutung⁶⁷.

Begreift man die möglichen Einordnungen und die Funktion von 3D-Stereofilm, kommt man zu dem Schluss das der 3D-Stereofilm nur als Stilmittel überleben kann. Ein Effekt hat nur einen kurzen Moment den Zuschauer von einem Eindruck zu überzeugen. Meist ist ein Effekt auch nur als stützendes Mittel einer Szene anzusehen. Häufig ist er aber ein Versuch eine Szene so aufzuwerten, das sie etwas besonderes vermittelt, was natürlich meist nicht gelingt da die Ausgewogenheit zwischen Inhalt und Form fehlt. Das besteht nur, wenn man alle Teile inhaltlich aufeinander abstimmt, das dramaturgisch und emotional passend sind. Dies ist nur Möglich wenn man die Intensität des verwendeten gestalterischen Mittels steuern kann. Die führt zu dem Schluss das ein 3D-Stereofilm nur ein Stilmitteln sein kann und auch nur als solches Gesehen und verwendet werden sollte.

Es sollte zwei Methoden unterschieden werden. Die beiden Methoden sind unter verschiedenen Vorraussetzungen und deren Folgen zu beobachten und zu beurteilen. Die ersten Methode setzt eine Entwicklung voraus, die den 3D-Stereofilm ganzheitlich sieht. Das bedeutet der Film nutzt die Technik nicht um einen bestimmten Effekt zu erzielen sondern als Stilmittel, der wiederum dem Film in seiner Art und Weise prägt und eine Richtung gibt.

Die andere Methode beschreibt genau das Gegenteil, nämlich einen zeitlich- und handlungsbegrenzenden Teil eines Moments im Film, mit dem Ziel den Zuschauer für diesen bestimmten begrenzten Bereich optisch aufzufangen.

5.5. Rückwirkungen der Stereoskopie auf tradierte kinematographische Codes

Die wesentlichen Punkte dazu sollen hier bezüglich der tangierten kinematographischen Codes aufgelistet und sortiert werden:

67 Vgl. David Bordwell Film Art: An Introduction, 2003, S. 164.

5.5.1 Photographisch/Bildkompositorisch:

- Kein seitlicher Anschnitt von Vordergrundobjekten (window-violation).
- Alle bekannten Einstellungsgrößen von Panorama bis Detail/Makro sind prinzipiell auch im S3D möglich. In den extremeren Einstellungen muss allerdings die Stereobasis deutlich jenseits des natürlichen Augenabstands verändert werden es kann zu übersteigerten Tiefenwahrnehmungen kommen (Gigantismus, Mikropsie).
- Große oder schnelle Tiefensprünge beim Umschnitt überfordern oft das visuelle System. Neue Herangehensweisen an Einstellungsgrößen/Perspektivkontraste müssen sich herausbilden.
- Sehr schnelle Kamerabewegungen lassen (derzeit) den L/R-Fusionsprozess im visuellen System zusammenbrechen, da aufgrund der Bewegungsunschärfe keine Kantenseparation mehr stattfinden kann.
- Lichtsetzung: Evtl. Trend zum kontrastreicheren Lichtsetzen als heute (durch Erzeugung von Lichtstaffelung in der Szene erhöht man generell jede Tiefenwirkung. Konsequentes Backlighting erhöht die Querdisparitätswahrnehmung des Zuschauers. Gut wahrnehmbare Texturen erhöhen die Stereowirkung dramatisch. Achtung: Überstrahlungen/Reflektionen in einer Szene verwischen die Disparitätswahrnehmung und zerstören unmittelbar den Tiefeneffekt in diesem Bildbereich).

5.5.2 Schauspiel

Schauspieler (aber auch Stuntmen) müssen nun auch bei mittleren Einstellungsgrößen mit erhöhter räumlicher Disziplin innerhalb der Stereo-Komfortzone agieren. Im Gegenzug bekommt die vom Schauspieler verkörperte Figur eine beeindruckende „Tiefe“ und „Präsenz“.

5.5.3 Ton

Antonio Lopes weist im Artikel „Stereoskopisches Video mit Wellenfeldsynthese-Audio“ im vorangegangenen Arbeitspapier Nr.5 darauf hin, dass je stärker der Zuschauer in das visuelle Geschehen eingebettet wird (insbesondere bei großen Leinwänden), die Notwendigkeit von 3D-Tonaufzeichnung und Tonproduktion egal in welcher technischen Ausprägung - immer notwendiger wird da es sonst bei der Filmbetrachtung schnell zu kontraproduktiven Wahrnehmungsdivergenzen zwischen visuellen und auditiven System kommen kann. In seinen Ausführungen weist er darauf hin, dass je nach Höhe der Leinwand insbesondere auch die vertikale Positionierung von sichtbaren Klangquellen berücksichtigt werden muss.

5.5.4 Montage

Ein Schnitt muss nun nicht nur im Sinne der klassischen Montage, an der richtigen Stelle gesetzt sein, sondern er darf in der Folgeeinstellung auch nicht aus der Tiefe fallen oder zu stark in eine neue Tiefe gesetzt sein. Weiche Übergänge oder eine stetige Steigerung oder Absenkung der Tiefe sind hierbei entscheidend. Durchgeführt kann dies entweder photographisch oder bildkompositorisch in einer Szene selbst oder durch nachträgliche Parallaxe-Anpassungen der Einstellungen vor dem Schnitt. Bei einer guten Tiefenplanung (s. Kap. 5.6 „Tiefenscript“) kann Letzteres meist entfallen.

5.5.5 Rückbezüge zum 2D-Film

Einen nativen 3D-Stereofilm im eigentlichen Sinne wird der Zuschauer derzeit oft noch gar nicht zu sehen bekommen sondern meist ein Mischprodukt basierend auf herkömmlichen Filmstilen, Bildsprachen, Dramaturgien und Stereo. Zum einen muss ein Film auch in einer 2D-Fassung erhältlich sein. Da sicherlich keine parallele 2D-Fassung produziert wird, eine monoskopische Rückkonvertierung aber technisch gar nicht so einfach zu bewerkstelligen ist, muss davon ausgegangen werden, dass bei vielen gestalterischen Möglichkeiten ein Kompromiss zu schließen ist. Der dramaturgische Einfluss dieser Kompromisse ist dabei schwer abzuschätzen.

5.6 Neue Gestaltungs-, Steuerungsmöglichkeiten der Stereografie

Hinsichtlich der Gestaltungsmöglichkeiten des stereografischen Codes, gibt es trotz aller syntaktischen Restriktionen bezüglich formaler filmischer Techniken einige Methoden diese dramaturgisch zu planen, zu gestalten und zu steuern.

Als wichtigstes Werkzeug gilt derzeit das sog. „Tiefenscript“. Mit diesem kann über den gesamten Film hinweg die Tiefe einzelner Szenen oder des gesamten Films geplant, begutachtet und bei bestimmten Szenen die für die Handlung wichtig sind auch dramaturgisch besonders hervorgehoben werden. Wichtige Ereignisse in der Handlung sollten eine besondere Tiefenwirkung haben, allerdings langsam aufgebaut und erst zum dramaturgischen Höhepunkt voll entfaltet werden um den Zuschauer bestmöglich darauf vorzubereiten. Der Effekt stützt dabei in symbiotischer Eintracht die Dramaturgie und umgekehrt⁶⁸.

In der Abbildung 8 ist beispielhaft ein Tiefenscript zu einem Animationsfilm abgebildet⁶⁹. Dabei repräsentiert die schwarze Linie die Leinwand, alles was über der Linie ist wird vor der Leinwand wahrgenommen, alles andere dahinter ist wirkt vermehrt in die Tiefe. Die rote Linie kann ein Objekt oder Darsteller sein. Das Orangene ist die Darstellung von unwichtigeren Elementen wie zum Beispiel fallende Blätter oder etwas das durch das Bild fliegt. Die blaue Linie ist eine Landschaft im Hintergrund, da-

68 Vgl. Mendiburu 3D Movie Making, 2009, S. 88.

69 Vgl. Gardner Perception and The Art of 3D Storytelling, 2009a, S. 10

zugehörend noch eine gestrichelte Linie die zum Beispiel einen Himmel beschreiben könnte. Die schwarzen vertikalen gestrichelten Linien sind die geplanten/gesetzten Szenenwechsel. Man beachte dabei, dass es - gemäß der stereoskopischen Montage-Empfehlungen - innerhalb der Szenen zu eher geringen oder aber rampenförmigen Tiefenveränderungen kommt und bei Szenenwechseln (mehrheitlich) zu Tiefen-Anschlüssen. Bemerkenswert im ersten Viertel ist auch die langsame Hinführung der Zuschauer zu stärkeren Tiefen, eine Augenpause zu Beginn der zweiten Hälfte (nach einem ersten dramatischen Höhepunkt) sowie die regelmäßige Rückkehr der roten Line (Hauptebene des Zuschauerinteresses) zur „Nulltiefe“ bei Szenenwechseln.

Ziel eines jedes Tiefenscripts ist es also, darzustellen mit welcher Tiefe eine stereoskopische Szene geplant ist und wie man den Tiefenverlauf dramaturgisch zum Inhalt anpasst.

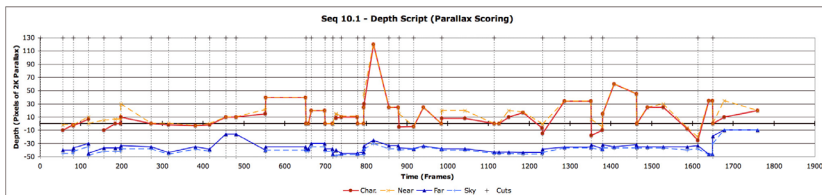


Abbildung 8: Tiefenscript von Brian Gardner

5.7 Dramaturgische Grundkonzepte des Stereofilms

5.7.1 Immersion

Spontan wird als dramaturgisches Grundkonzept der Stereographie meist eine gesteigerte Tiefenwirkung für den Zuschauer proklamiert (im Sinne einer Verlängerung der klassischen Mittel der Mise en scène). Die passenden Genres und Anwendungsfälle für diesen Ansatz sind eher selten. Tanz- und Ballettfilme bilden derzeit die rühmliche Ausnahme.

Wie schon mehrfach ausgeführt, ist das eigentliche zentrale dramaturgische Grundkonzept des Stereofilms die filmische Immersion des Zuschauers. Schon die klassischen kinematographischen Codes versuchen durchweg dieses Ziel zu erreichen, mit der Stereoskopie wird diesem Ansinnen ein äußerst wirksames Stilmittel zur Verfügung gestellt. Ein kurzes Beispiel zur Verdeutlichung ist der klassische „Establishing-Shot“, der den Zuschauer visuell in eine Szene einführen soll. Seine Wirkung kann mit Stereo problemlos bis zur Immersion verstärkt werden.

5.7.2 Leinwand als imaginäre Grenze

Im 2D-Film stellte die Leinwand immer eine imaginäre Grenze zwischen dem Zuschauer und der Filmwelt dar. Er konnte zwar emotional in als Geschehen „hineinge-

zogen“ werden, die Filmwelt kam aber nie zu ihm heraus. Mit der stereoskopischen Möglichkeit der Off-Screen-Objekte war dies nun technisch machbar. Leider wurde diese neue grenzüberschreitende gestalterische Möglichkeit von den frühen Stereofilmen - oft ohne jegliche dramaturgische Konnotation - so begeistert ausgenutzt, dass sie mittelfristig ihre Wirkung verlor (abgesehen von der körperlichen Anstrengung des Sehapparates die überproportionale Off-Screen-Objekte erfordern). Es kam durch diese Effekte im Gegenteil eher zu einem Abgewöhnung-Effekt. „Gimmicks ziehen den Zuschauer aus der 3-D-Erfahrung heraus“, sagte dazu ROBERT NEUMANN 3D-Supervisor für den 3D-Animationsfilm BOLT. Seiner Meinung nach zerstören sie die mühsam aufgebaute Immersion nachhaltig⁷⁰.

Desweiteren ist diese Grenzüberschreitung psychologisch natürlich heikel, da sie vom Zuschauer auch als „Angriff der Filmwelt“ interpretiert werden kann und er das Vertrauen zum Plot verliert (Anm.: womöglich noch heikler beim 3D-TV, da hier im Gegensatz zum Kino tatsächlich ein Einbruch in die Privatsphäre stattfindet). Es ist daher in einigen aktuellen Stereofilmen ein eher konservatives Verhältnis zu Off-Screen-Elementen auszumachen. Wie in der vorherigen Abbildung 8 beispielsweise zu erkennen ist, gibt es in diesem Film nur wenige Szenen mit sehr starker Ausrichtung zum Zuschauer hin (angezeigt durch den starken Ausschlag der roten Linie). An diesen „Plotpoints“ wird der Stereoeffekt nur punktuell verstärkt, weil etwas emotional Wichtiges passiert. Jenseits davon wird mit off-Screen-Anteilen eher moderat umgegangen.

Langfristig interessant wäre nun, einen Stereocode herauszubilden, der den Zuschauer an die Leinwand als imaginäre Grenze gewöhnt und ihn Überschreitungen als tieferes Verständnis des Gezeigten wahrnehmen lässt im Sinne von Inszenierungen zum Zuschauer hin. Dies geht womöglich mit einer bewussten Einschränkung der technisch-machbaren Tiefenmöglichkeiten einher – es könnte aber die Immersion und damit die Wirkung der Filmbotschaft deutlich verbessern.

5.7.3 Korrelation von Räumlicher und Emotionaler Distanz

Die räumliche Distanz zum Filmgeschehen kann und darf im 3D-Stereofilm verlängert oder überwunden werden, die Stereoskopie macht dies möglich. Sie korreliert aber (ähnlich wie weite oder nahe Kameraeinstellungen) mit der emotionalen Distanz des Zuschauers zum Geschehen. Grundsätzlich und losgelöst von der Off-Screen-Thematik ergibt sich für die Filmschaffenden also auch eine neue Möglichkeit emotionale Distanz aufzubauen.

Bestimmte Momente mit inhaltlicher, dramaturgischer und emotionaler Tiefe ergeben den Grund dafür die Distanz zu verringern. Welche dramaturgischen Vorteile sich dadurch ergeben liegt auf der Hand: der Effekt bindet sich an einen emotionalen Höhepunkt, was dazu führt dass dessen Darstellung und Aufnahme intensiviert wird. Um-

70 Vgl. Merschmann Wenn die Gitarre aus der Leinwand ragt.

gekehrt kann (wie z.B. mit einer klassischen Vogelperspektive durch die Kamera) im umgekehrten dramaturgischen Ansatz mit erhöhter (Tiefen-)Distanz die emotionale Distanz nochmals verstärkt werden

Interessant wären letztendlich erprobte Konzepte diese Korrelation in der Intensität zu steuern und zum Gesehenen anzupassen. Ein spannendes Feld für zukünftige S3D-Filmschaffende.

5.8 Analyse ausgewählter Produktionsbeispiele 2009

Kinobeispiele zum Zeitpunkt 2009; zwei Animationsfilme und zwei Spielfilme (mit hohem Buescreen/CGI-Anteil). Interessanterweise jeweils einer im Genre zunächst nicht stereoskopisch gedreht.

OBEN

Man darf behaupten dass mit OBEN⁷¹ einer der ersten richtigen Animations 3D-Stereofilm erstellt worden ist, der durch eine konsequente stilbildende Verwendung des Stereoeffekts einen eigenständigen Filmstil kreiert hat. In OBEN wurde Stereo zum ersten Mal nicht als Effekt genutzt, sondern konsequent als Mittel zur dramaturgischen Veredelung des Films. 3D-Stereo wurde somit Teil eines Filmprozesses, der es nicht mehr darauf angelegte ständig Objekte „off-Screen“ darzustellen, sondern eine Tiefendramaturgie zu entwickeln und diese der Dramaturgie des Plots zu unterwerfen. Man darf damit behaupten, dass mit OBEN einer der ersten konsequenten Stereo 3D-Animationsfilme erstellt wurde, der die Erwartungshaltung der Zuschauer an solche Filme veränderte und prägte.

GFORCE

Die Produzenten des Animationsfilms GFORCE hatten sich zur Aufgabe gestellt, einige Elemente des Films in 3D-Stereo zu produzieren. Es wurde also in einigen Szenen der stereoskopische Effekt als verstärkendes „Gimmick“ hinzugefügt. Es stellte sich dabei heraus, dass nicht alle tradierten gestalterischen Codes des Animationsfilms mit der Stereoskopie vereinbar waren. Der Einsatz von starker Tiefenunschärfe als grundsätzlicher Overall-Look des Films führte z.B. dazu, dass die stereoskopische Tiefenwirkung immer wieder spontan entschärft wird.

ALICE IN WONDERLAND

Obwohl nicht in S3D aufgenommen, wurde Regisseur TIM BURTON wohl relativ schnell klar, dass Stereo als grundlegendes Stilmittel der eher psychedelisch angelegten Handlung nochmals einen deutlichen dramaturgischen Schub verleihen kann. Mit finanziell und zeitlich/handwerklich hohem Aufwand wurde Schritt-für-Schritt aus dem monoskopischen Filmmaterial eine 3D-Stereo Version geplant und erstellt.

71 Pixar Animation Studios/Walt Disney, Originaltitel: Up, USA 2009, Regie: Pete Docter & Bob Peterson, Drehbuch: Bob Peterson, Musik: Michael Giacchino

Im eigentlichen Wonderland konnte die Tiefenwirkung natürlich problemlos überhöht dargestellt werden, dies wird bei Off-Screen-Objekten aber nur an dramaturgisch sinnvollen Stellen eingesetzt. Mehrheitlich wird auch in diesem Film auf Immersion gebaut. Der Film bietet einige gute Beispiele, wie Tiefe nicht zur Verstärkung des Raumes eingesetzt wird sondern auch zur Steigerung der (Volumen-) Präsenz einiger Charaktere. Dieses Stilmittel wird derzeit noch viel zu oft vergessen. Eine charmante „Versöhnungsgeste“ für die durchgeführten Durchbrechungen der Leinwand bietet der Abspann des Filmes, in dem BURTON langsam Blumen über den imaginären Bildrand hinaus in den Zuschauerraum wachsen lässt.

AVATAR

Bei Produktion von AVATAR lag der Fokus auf dem gesamten S3D-Produktionsprozess. Dieser wurde erstmalig in einem professionellen Großrahmen durchgeführt. Das Konzept einen 3D-Stereofilm zu produzieren, wurden in alle Filmentwicklungsphasen mit eingeplant. Aufgrund des hohen Anteils an Greenscreen-Aufnahmen konnten die letztendliche Tiefe der fertigen Bilder problemlos nach dramaturgischen Vorgaben „durchgestylt“ werden und war dadurch von technischen Aufnahmebegrenzungen weitestgehend befreit. Dadurch konnte die stereoskopische Intensität sehr gut an den dramaturgischen Aufbau angepasst werden. Tendenziell gibt es in AVATAR viele weite Realkamera-Einstellungen, die den Raum vor der Leinwand genauso nutzbar machen wie den Raum in und hinter der Leinwand.

Auffallend an diesem Film ist aber auch der zurückhaltende Einsatz von Stereotiefe. Um dafür eine Erklärung geben zu können, ist es notwendig zu verstehen, was für eine Bedeutung die einzelnen Szenen haben. CAMERON wählte dafür bewusst Szenen, die laut Drehbuch (auf dem fernen Planeten Pandora) einem flachen Realismus entsprachen und stellte diesen Realismus folglich auch so dar. Er bildete dadurch einen immersiven Filmstil, der den gesamten Film beeinflusst. Die Gründe warum CAMERON - der zu diesem Zeitpunkt schon viel Stereoerfahrung gesammelt hatte - generell die stereoskopische Wirkung reduzierte könnten angesichts des finanziellen Volumens dieses Films sowohl von dramaturgischer wie auch von strategischer Natur sein:

- Der Tiefeneffekt sollte sich nicht so schnell abnutzen, damit er an passender Stelle umso dringlicher wirkt.
- Eine große Anzahl an Zuschauern erleben zum ersten Mal einen 3D-Stereofilm.
- Die Wahrnehmung der Zuschauer für den 3D-Stereofilm musste sich damals erst noch entwickeln.
- Der Zuschauer soll sich bei einem 2 Stunden Film nicht übermäßig angestrengt fühlen.

6. Kritiken am 3D-Stereofilm

S3D-Produktionen kritisch zu hinterfragen ist für dessen Entwicklung durchaus notwendig. Die Vergangenheit zeigt deutlich auf, dass der Hype um S3D immer zwischenzeitliches Potential hat aufblitzen lassen, dieses Potential aber nie wirklich in einen echten Paradigmenwechsel hat einleiten können.

6.1 Stereofilme als Zuschauermagnet

S3D-Filme werden heutzutage meist in Form von Blockbustern⁷² realisiert. Tatsächlich liegt der Fokus an den Kinokassen, es sollen möglichst viele Zuschauer ins Kino gelockt werden. Wöchentlich starten S3D-Produktionen in den Kinos und die Zuschauer sollen möglichst viele davon sehen und die Extrakosten der Produktionen für S3D-Filmen begleichen.

Sicherlich stehen die Einnahmen von S3D-Filmen in Vordergrund und diese fallen bekanntlich höher aus als bei einer herkömmliche 2D-Produktion⁷³. Der Zuschauermagnet mit dem Prädikat „3D“ liefert den Produzenten eine echte Alternative, kostspielige Filmproduktionen effektiv zu vermarkten. Der Zuschauer verliert jedoch mit der Zeit seine Naivität - die Folge, vor allem bei den erhöhten Preisforderungen der Kinobetreiber und Verleiher, besonders bei mittelmäßigem Inhalt, könnte zu verminderter Akzeptanz jenes Hoffnungsträger werden, den die Filmindustrie seit Jahren gesucht hat. Die Gefahr besteht das "3D" nur als Marketinginstrument verbrannt wird. Eine Studie von PricewaterhouseCoopers⁷⁴ ergründet, dass der Zuschauer durchaus an dreidimensionalen Inhalten interessiert ist, aber dass nach der anfänglichen Neugier auf S3D-Filme der Zuschauer dauerhaft nicht bereit sein könnte die erhöhten Ticketaufschläge zu bezahlen. Hier müssten die Kinobetreiber und Verleiher bereit sein in irgendeiner Art und Weise dem Zuschauer entgegen zu kommen oder gar komplett auf den Aufschlag zu verzichten. Es darf jedoch nicht vergessen werden, dass der Zuschauermagnet „3D“ bei vielen Produktionen nur scheinbar als Erfolg bezeichnet wird, falls man den 3D-Zuschlag abziehen würde, bleibt nicht mehr so viel vom eigentlichen Erfolg übrig.

72 <http://de.wikipedia.org/wiki/Blockbuster>

73 <http://www.bluray-3d.de/3d-filme-erzielen-in-den-usa-hoehere-einnahmen.html>

74 Studie des PwC-3D Tipping Points Loom 2009

6.2 Brillenproblematik

Die Brillenproblematik ist eine technische Hürde, da neue Entwicklungen, die es möglich machen mehreren Personen ein brillenloses 3D-Erlebnis zu gewährleisten, aktuell noch ausbleiben. Die Zuschauerakzeptanz von Brillen, die zweifelsfrei notwendig sind, muss dahingehen näher beleuchtet werden. Auch ökonomisch betrachtet könnte die Brille zu Schwierigkeiten führen, da Kinos auch hier einen Zuschlag für 3D-Filme verlangen. Fernsehlösungen, dabei ist es egal ob als Plasma oder LCD, sind ebenso betroffen, da zu deren Nutzung auch Brillen benötigt werden. Dabei sind jedoch zusätzlich ziemliche Preisunterschiede zwischen Systemen mit Shutter- und Polarbrillen, besonders bei der erweiterten Peripherie, vorhanden.

Eine Studie der PRIME-Forschungsgruppe untersuchte diese Problematik und erforschte die Akzeptanz bei der Nutzung von 3D-Brille. 54,4 % der Befragten gaben an, dass sie keine nennenswerten Probleme mit dem 3D-Erlebnis aufweisen, wenn sie dazu eine 3D-Brille nutzten⁷⁵. Weitere 22 % gaben an, dass es sie durch die 3D-Brille zu keinem verschlechterten Seherlebnis kommt wenn sie eine 3D-Brille nutzten. Allgemein kann man davon ausgehen, dass sich mit der Häufigkeit der Nutzung diese Diskrepanz verringert.

Brillenproblematik und das Auge

Die Frage die sich hierbei stellt, ob es möglich ist, dass sich das Auge auf verstärkte Divergenzbildung oder generell beim Fusionsprozess, eine Gewöhnung stattfindet. Immerhin ist das Auge nur ein Muskel, der wie jeder andere im Körper nur trainiert werden könnte. So haben sich 3D-Trailer vor dem eigentlichen Film als sehr wirksam für die „Eingewöhnungsphase“ erweisen. In der „Great Debate“ im Slate Magazin wird genau diese Eventualität aufgegriffen. Hierbei wurde beispielhaft diskutiert, dass es durch häufigen Konsum von S3D-Filmen zu einer „subjektiven“ Eingewöhnung kommt. Schlecht produzierten S3D-Filmen könnten dadurch auch subjektiv weniger Schwierigkeiten bereiten und es würde zu einer Art der De-Sensibilisierung⁷⁶ des Auges führen. Vielleicht ist dies auch im Moment nur eine Übergangsphase, da Grenzen der Wahrnehmung immer Zeit zur Überwindung benötigen, wie es auch beispielhaft bei schnellen Schnittfolgen zunächst immer eine Gewöhnungsphase gebraucht hat.

6.3 Irritationen beim S3D Seherlebniss

Wie kann ein Film der flimmert, Kopfschmerzen bereitet, anstrengend ist und irgendwie immer zu dunkel wirkt überhaupt ernst genommen werden. Eigentlich hat man als Zuschauer mehr zu verlangen, aber dieser ist sogar bereit mehr für einen „flim-

75 Media Perspektiven": 3D Kino im Urteil des Publikums, 2009

76 <http://www.slate.com/id/2265255/entry/2265342/>

mernden“ Film zu bezahlen. Diese skeptischen Ausführungen stammen von Roger Ebert, ein hochangesehener amerikanischer Filmkritiker. Er beschreibt seine Kritik außerdem mit den Sehgewohnheiten des menschlichen Auges und dem Effekt der Schärfentiefe. Alles worauf der Zuschauer seinen Blick richtet ist scharf, alles was davor oder dahinter, ist unscharf. Die Sehgewohnheiten sind also mit der wahrnehmbaren Distanz verbunden. S3D löst diese Verbindung von Distanz und Schärfe auf. Der Effekt der Schärfentiefe wird bei S3D-Filmen nicht verworfen, allerdings kann man nun nicht mehr genau wissen worauf der Zuschauer gerade blickt, somit ist die Zuordnung von Distanz und Schärfe nicht mehr möglich⁷⁷. Außerdem ist seiner Meinung nach ist dies auch der Grund, weshalb sich S3D niemals so wirklich durchsetzen wird.

6.4 Inhaltliche Einsätze von S3D

Sieht man davon ab, wie S3D oder der reine Effekt historisch eingesetzt worden ist oder er in der Pre-Avatar Zeit verstanden worden ist, dann erkennt man heute das der 3D-Effekt nicht mehr als Effekt verstanden werden soll, sondern als Einstieg in ein immersives Erlebnis, dass die Leinwand nicht mehr als Grenze sieht, sondern als eine Fülle von Möglichkeiten, davor und dahinter. Jedoch ist das auch genau ein Kritikpunkt an S3D-Filmen. Die Leinwand ist genau das, eine Grenze, ein Blick durch eine Kamera mit dem „Versprechen auf Mehr“⁷⁸. Ein Rahmen der dem Filmemacher seinen Spielraum bietet dem Zuschauer Informationen weiterzuleiten und diese Informationen nach seinen Vorstellungen zu gestalten. Das Verlassen der Kadrierung, das „Versprechen auf Mehr“ wird dadurch ihrer Bedeutung beraubt, die filmische Realität wird aufgebrochen und dem Zuschauer wird bewusst, dass es sich um einen Täuschung handelt, es zeigt Grenzen auf die vorher nicht da waren. Die allgemeine Aussage von 3D-Filmen ist die der „neuen Realität“ oder den Erhalt eines ganzen neuen „filmischen Erlebnisses“, doch ist es das wirklich? Der Zuschauer fühlt das die Realität irgendwie filmisch eine andere geworden ist, sie wirkt plastisch und unecht, vielleicht mag das für gewöhnliche Blockbuster-Filme reichen, aber Inhalte mit hohem Anspruch und tiefen Geschichten, Bildern die auch gerne Informationen offen lassen, sollte das „Verlangen nach Mehr“ bei Zuschauererzeugen und keine flache Geschichte mit perfekten 3D-Effekten „ausfüllen“ und „abrunden“⁷⁹.

77 http://blogs.suntimes.com/ebert/2011/01/post_4.html

78 Bayer, F.: Künstliche Weiten: Ein Kommentar zu James Camerons Avatar

79 Bayer, F.: Künstliche Weiten: Ein Kommentar zu James Camerons Avatar

7 Fazit

Auch wenn der Themenbereich der Stereoskopie schon seit dem letzten Jahrhundert von der Filmindustrie immer nur dann in den Fokus gerückt wurde, wenn sich Krisen angedeutet haben, steht der jetzige Hype nicht unter den besten Voraussetzungen, um eine Innovation bis zu Ihrer Perfektion zu erforschen. Momentan herrscht die selbe Situation wie in den 1950er Jahren, auch wenn damals das Bewusstsein um die Qualität der Filme eine andere war und die Erwartungshaltung der Zuschauer sich ebenfalls gewandelt hat. Was darauf zu begründen ist, das der Zuschauer immer mehr Perfektion sehen möchte, welche er durch seine angeeigneten Erfahrungswerte erlangt hat. Er möchte gutes Kino mit starken Bildern, perfekte Zusammensätze und schöne Geschichten sehen und diese auch erleben. Dabei steht für ihn nicht das Wie und das Was im Vordergrund, sondern das gesamte Filmerlebnis. Da 3D-Stereo Technik ein Zugpferd für viele zukünftige Produktionen darstellt, verbessert dies allerdings die Qualität der Produktionen nicht sonderlich. Es kann den Film nur in der bisherigen Form, in bestimmten Szenen, Sequenzen oder als Teil eines ganzen formalen Filmsystems, das sich dem Inhalt unterordnet, unterstützen.

Die Form beeinflusst den Inhalt, sie erzeugt das filmische Gebilde, dadurch bedingt es die Möglichkeiten den Inhalt darin zu konstruieren. Ein Teil dieses Filmgebildes ist die Stereoskopie, es schränkt die Erzeugung des Inhalts nicht ein, sondern verleiht dieser seine ganz eigene Charakteristika. Diese Ausprägung sorgt für eine Veränderung der Bildsprache. Sie wird Teil der gesamten Kette an Produktionsabläufen. Jede Szene muss auf emotionale Tiefe hinterfragt werden, um anschließend diese auf die räumliche Tiefe angepasst zu werden. Dadurch bildet sich eine neue, erweiterte Form der Dramaturgie heraus. Das Kompositum von Handlungsabläufen wird direkt mit der räumlichen Tiefe verbunden und bildet somit ein ständig anpassenden Moment für den Film. Dabei muss allerdings auch ein Maß gefunden werden, wie weit ein 3D-Stereofilm die Anzahl und die Stärke der Stereoansichten einschränkt.

Dies hängt zum einen davon ab wie lange die Dauer des Film ist und zum anderen wie stark eine Szene damit unterstützt werden soll. Wie bei jedem Film im Vorfeld auch ein bestimmter Klangteppich oder eine bestimmte Musikauswahl getroffen wird, sollte das stereoskopische Mittel in seinen Ausprägungen ausgewählt werden. Dies gelingt, wenn das filmische Gebilde und das inhaltliche Konstrukt miteinander verbunden werden. Es sollte zusätzlich zu dem musikalischen "Klangteppich", somit ein visueller stereoskopischer "Klangteppich" entstehen.

Das ausgerufenen Ziel der Stereoskopie ist, die Verstärkung des Bild um ein immersives Erlebnis im Kino zu erzeugen. Die Stereoskopie ist dabei nur ein Teil einer ganzen Prozesskette an Vorgängen, die dazu dienen das zu gewährleisten. Emotionen sind ein weiterer wichtiger Teil eines immersiven Erlebnisses. Dabei kann die Stereoskopie in Verbindung von Bild und Ton, als Transportkanal für Emotionen angesehen werden.

Sie bildet die aktive Teilhabe an Film und seinem Inhalt heraus, koppelt Teile heraus und sorgt in bestimmten Szenen auch für Ruhe. Dabei spielt die Bildsprache eine große Rolle. Die Wahl des Format, der Kardrierung oder Farbzusammenhängen beeinflusst den Kanal ebenfalls. Die Überwindung der räumlichen Distanz ist allerdings die Stereoskopie vorbehalten. Herkömmliche Filmtechniken erzeugen nur ein Klima, dass durch das zweidimensionale Abbild erahnt werden kann. 3D-Stereofilm hingegen bringen das Abbild zum Zuschauer hin oder als anderer Ansatz den Zuschauer hin zum Abbild. Weiterhin ist es notwendig, dass ein Film zu einem "synästhetischen Erlebnis" wird. Dies generiert neben der visuellen auch eine weitere Tiefenebene, das aus einem mentalen Prozess des "Hineinfühlens" entsteht.

Weiterhin wurde die Fragestellung aufgeworfen unter welchen Gesichtspunkten eine Grammatik aus dem 3D-Stereofilm heraus gebildet werden kann. Wird eine Teilung zugelassen, in der die Grammatik in technische Möglichkeiten und gestalterische Möglichkeiten geteilt wird, ergibt sich, dass die Technik im Grunde nicht verändert, aber die gestalterischen Möglichkeiten stetig einer Erweiterung unterliegen werden. Es werden immer Restriktionen und Dinge geben, die durch die Stereoskopie nur beschränkt oder gar nicht funktionieren werden. Sie bilden die technischen Codes mit ihren Subcodes. Die gestalterischen Möglichkeiten beschreiben sich hingegen auf dessen Wirkung. Das bedeutet, auch die Verzahnung mit herkömmlichen Codes wie Fotografie und Montage ergeben sich nur mit der Erweiterung der Wirkungsmethodik. Die Stereoskopie hat sich in den letzten Jahren als Nische langsam entwickelt und eine gewisse Reife entwickelt, die von der Masse an Zuschauern angenommen wird. Es gilt dies auszunutzen und Inhalte zu entwickeln, welche es auch wert sind eine stereoskopische Untermalung zu bekommen. Mit AVATAR wurde hier der Grundstein gelegt und jeder anspruchsvolle Filmemacher, sollte sich mit dieser neu aufgelegten Thematik vertraut machen. Für eine Erforschung der Möglichkeiten müssen weitere Experimente durchgeführt werden, um den 3D-Stereofilm in Zukunft zu erweitern. Die Entwicklung des Films bzw. des 3D-Stereofilms wird keinen Stillstand erfahren und sich stetig weiter entwickeln.

Literaturverzeichnis

- Bazin, André; Hartmut Bitomsky, Harun Farocki und Ekkehard Kaemmerling (Hrsg.): Was ist Kino : Bausteine zur Theorie des Films, Verlag M. DuMont Schauberg, 1975
- Bayer, Florian: Künstliche Weiten: Ein Kommentar zu James Camerons Avatar. 2010
- Borwell, David; Rost, Andreas (Hrsg.): Visual Style in Cinema - Vier Kapitel Filmgeschichte, 3. Auflage. Frankfurt am Main: Verlag der Autoren, 2006
- Century Fox PresseService: Bericht über Einspielergebnisse, September 2009, <http://foxfilm.demedianetworx.de> URL: <http://foxfilm.demedianetworx.de> [Stand:25.02.10]
- Claudia Wegner, Jesko Jockenhövel: 3D- Kino im Urteil des Publikums, 9 2009, Quantitative Studie zur Nutzung und Akzeptanz
- Curtis, Robin: Zusammenfassung von der Tagung: Empfindungsräume - Zur synästhetischen Wahrnehmungsformen, in: Montage-AV, 2009 URL: <http://www.nachdemfilm.de/report/synaest.html> [Stand:25.02.10], Beitrag: Immersiven Räumlichkeit von Filmen
- Chicago Sun-times blog, Roger Ebert, Why 3D doesn't work and never will. Case closed, URL: http://blogs.suntimes.com/ebert/2011/01/post_4.html [Stand: 10.07.2011]
- David Bordwell, Kristin Thompson; Freitag, Chris (Hrsg.): Film Art: An Introduction, McGraw-Hill, 2003
- Faulstich, Werner; Fink, Wilhelm (Hrsg.): Grundkurs Filmanalyse, 2. Auflage. Paderborn: Wilhelm Film Verlag, 2008
- Frank, Anja: Stereoskopische High Definition Produktion, Diplomarbeit, HFU Hochschule Furtwangen University, SS 2007
- Gardner, Brian: Perception and The Art of 3D Storytelling, 2009, Creative Cow Magazine URL: <http://magazine.creativecow.net/article/perception-and-the-art-of-3d-storytelling> [Stand:25.02.10] Gardner, Brian: Stereoscopic 3D Issue, in: Creative COW Magazin, Juni 2009, 52
- Goldstein, E. Bruce; Irtel, Hans (Hrsg.): Wahrnehmungspsychologie - Der Grundkurs, 7. Auflage. Spektrum Akademischer Verlag der Autoren, 2008
- Guski, Rainer; Guski, Rainer (Hrsg.): Wahrnehmen - Ein Lehrbuch, Verlag W. Kohlhammer, 1996
- Hofmeyer, Frank: Analyse und prototypische Umsetzung eines Prozesses zur stereoskopischen Fusion von CG-Environments und Real-Footage Elementen, Diplomarbeit, HFU Hochschule Furtwangen University, WS 2007/2008
- Jockenhövel, Claudia Wegner & Jesko: 3D Kino im Urteil des Publikums, 09 2009, 8, 3D Kino im Urteil des Publikums
- Katz, Steven D.; Utecht, Harald (Hrsg.): Shot by Shot: Die richtige Einstellung, Zweitausendeins, 2004

- Knut Hickethier, Hartmut Winkler; Bohn, Sigma (Hrsg.): Filmwahrnehmung, Sigma Medienwissenschaft, 1990
- Kritzenberger, Huberta: Multimediale und interaktive Lernräume, Oldenbourg Wissenschaftsverlag, 2004
- Kuhn, Gerhard: Stereo-Fotografie und Raumbild-Projektion, vfv-Verlag Gilching, 1999
- Lipton, Lenny: Foundations of the stereoscopic cinema, Van Nostrand Reinhold Company, 1982
- Marks, Gerold: Digitale Leinwand - Die Digitalisierung des Kinos, 10.2009, URL: http://www.chip.de/news/3D-im-Kino-Zuschauer-wollen-laut-Studie-mehr-Filme_38394150.html [Stand:25.02.10]
- Marschall, Susanne: Farbe im Kino, Schüren Verlag GmbH, 2005
- Mendiburu, Bernard: 3D Movie Making, Elsevier - Focal Press, 2009
- Merschmann, Helmut: Wenn die Gitarre aus der Leinwand ragt, URL: http://www.epd-film.de/themen_66828.php [Stand:25.02.10]
- Metz, Christian: Semiologie des Films, München (Fink), 1972
- Mikunda, Christian; Bohn, Sigma (Hrsg.): Filmwahrnehmung, Knut Hickethier, Hartmut Winkler im Verlag der Sigma Medienwissenschaft, 1990
- Monaco, James; Bock, Hans-Michael (Hrsg.): Film Verstehen, 9. Auflage. Rowohlt Taschenbuch Verlag W. Kohlhammer, 2007
- Ohler, Peter; Bohn, Sigma (Hrsg.): Filmwahrnehmung, Knut Hickethier, Hartmut Winkler im Verlag der Sigma Medienwissenschaft, 1990
- Schnell, Ralf : Medienästetik - Zu Geschichte und Theorie audiovisueller Wahrnehmungsformen, J.B. Metzlersche Verlagsbuchhandlung und Carl Ernst Poeschel Verlag GmbH, 2000
- Schumacher, Eric: Analyse der Gestaltungsmöglichkeiten von stereoskopischem unkomprimiertem HD-Videofilm in Kombination mit Audio-Wellenfeldsynthese auf Immersionsbasis, Diplomarbeit, HFU Hochschule Furtwangen, SS 2008
- Smith, Murray; Tobias Ebbrecht, Thomas Schick und (Hrsg.): Emotionen - Empathie - Figur: Spielformen der Filmwahrnehmung, VISTAS Verlag GmbH, 2008
- Steinborn, Bion: Das audiovisuelle Bild und das Wirkliche, Frankfurt/M Fischer, 1996
- Slate-Magazin - The Great 3-D Debate, URL: <http://www.slate.com/id/2265255/entry/2265342/> [Stand: 10.07.2011]
- Voss, Robin Curtis & Christiane: Immersion, in: Montage-AV 2 2008
- Zone, Ray; Kentucky, The University of (Hrsg.): Stereoskopie Cinema and the Origins of 3-D Film, The University of Kentucky, 2007
- The Big Question: Are we heading towards a 3D screen future?, Rob Sharp Interview: Jeffrey Katzenberg, URL: <http://www.independent.co.uk/arts-entertainment/films/features/the-big-question-are-we-heading-towards-a-3d-screen-future-1868356.html> [Stand: 21.05.2010]

Autoren

NIKOLAUS HOTTONG



Prof. Nikolaus Hottong lehrt und forscht an der Hochschule Furtwangen Fakultät Digitale Medien - Fachgebiet Medientechnologien.

ERIC SCHUMACHER



Eric Schumacher, Absolvent des Master of Science im Studiengang Computer Science in Media an der Hochschule Furtwangen.

Interessen: Digital Cinematography, Postproduction, Fotografie, Musik, Reisen

FRANK HOFMEYER



Frank Hofmeyer, Absolvent des Master of Science im Studiengang Computer Science in Media an der Hochschule Furtwangen. Derzeit wissenschaftlicher Mitarbeiter am Institut für Medientechnik an der TU Illmenau.

DIRK HENSEL



Dirk Hensel studierte Medieninformatik an der HFU und ist derzeit im Bereich Marketing/Digitale Medien bei der RENA GmbH in Gütenbach tätig.

MARTIN MAURER



Bachelorstudium der Online Medien an der Hochschule Furtwangen University und Digital Media an der RMIT University in Melbourne, Australien. Derzeit Masterstudium der Medieninformatik an der Beuth Hochschule für Technik in Berlin.

PAUL WALTER



Studierte Online Medien an der Hochschule Furtwangen und absolviert im Moment ein Masterstudium der Digitalen Medien an der Hochschule Furtwangen.

